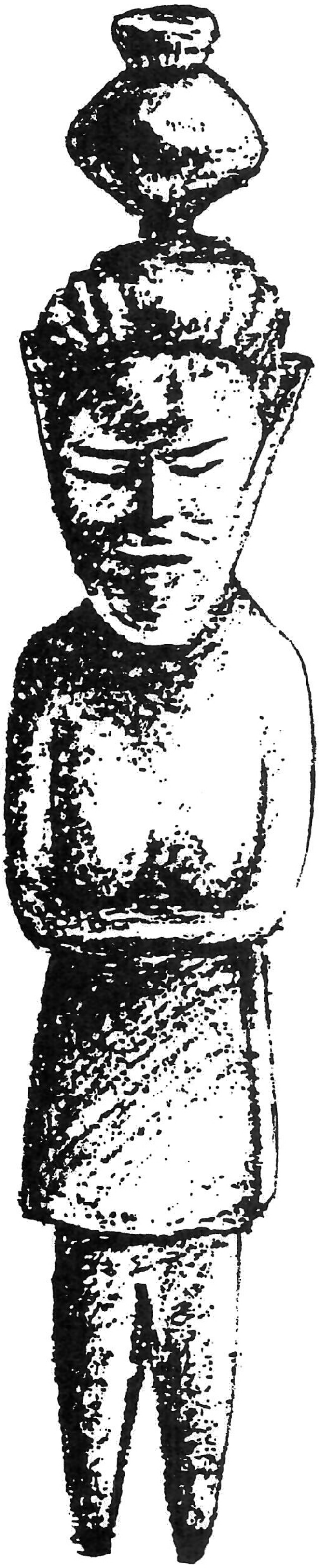
MOTHER TONGUE

JOURNAL OF THE ASSOCIATION FOR THE STUDY OF LANGUAGE IN PREHISTORY

Issue XIX • 2014



CONTENTS

1 The 200,000-Year Evolution of *Homo sapiens sapiens* Language and Myth Families based on mtDNA Phylotree Archaeogenetics: A Thought Experiment
James B. Harrod

87 The Kinship Term *KUKU ~ KOKO ~ KAKA* in the American Indigenous Languages, the Amerind Hypothesis, and the Dravidian Kinship System
Section I: Linguistic Study
Alain Matthey de l'Etang & Pierre J. Bancel

155 Indo-European and Dravidian: Some Considerations
Stephan Hillyer Levitt

181 Burushaski and the Western Dene-Caucasian Language Family: Genetic and Cultural Linguistic Links
John D. Bengtson

Book Reviews

223 E.J. Michael Witzel: *The Origins of the World's Mythologies*
Reviewed by *Václav Blažek*

229 Michael Fortescue: *Language Relations across Bering Strait: Reappraising the Archaeological and Linguistic Evidence.*
Reviewed by Peter A. Michalove

233 News & Notices: Association for the Study of Language in Prehistory (ASLIP)

MOTHER TONGUE Issue XIX · 2014
Journal of the Association for the Study of Language in Prehistory (ASLIP)

For subscription and membership information contact:

Secretary-Treasurer: Michael T. Lewis
20 Duane Avenue
West Newton, MA 02465
U.S.A.
lewismtc@rcn.com
Tel. 617-964-0978

ASLIP **membership: $35.00 USD,** yearly
Lifetime membership: $500.00 USD
For electronic payment see ***PayPal*** link at http://aslip.org/

For other information about ASLIP, or to inform ASLIP of important news, new scientific developments, and media relations, please contact:

Information Officer: Jonathan Sherman Morris
São Paulo, Brazil

jonathanmorris1964@gmail.com Tel: 5511-31512667

For submission of articles, notes, book reviews, or letters to MOTHER TONGUE, please contact:

Editor: John D. Bengtson
Savage, MN U.S.A.
palaeojdb@hotmail.com
Tel. 612-839-3649

For any other correspondence with ASLIP officers and advisors, see the contact information on the **inside front cover** of this volume.

ISSN 1087-0326 **ASLIP** Homepage: http://aslip.org

MOTHER TONGUE
Journal of the Association for the Study of Language in Prehistory • **Issue XIX • 2014**

The 200,000-Year Evolution of *Homo sapiens sapiens* Language and Myth Families based on the mtDNA Phylotree, Fossil mtDNA and Archaeology: A Thought Experiment

James B. Harrod
Director, Center for Research on the Origins of Art and Religion

Abstract. A meta-pattern-analysis of the mitochondrial DNA phylotree and current distribution of language families indicates that over the last 200,000 years there are robust correspondences between mtDNA haplogroups and language macrofamilies. This study is a thought experiment, a top-down derivation of the *Homo sapiens sapiens* ('Proto-Human', 'Proto-World') language phylotree, which can be tested against bottom-up prehistoric linguistic reconstructions. It establishes a relative chronology for dating the emergence and branching of the global array of language macrofamilies. The language phylotree is crosschecked against archaeological data and fossil mtDNA studies, which support many of the correlations. Results imply L3M and N dispersed out-of-Africa at around 80,000 years ago with both Afrasian and Nile-Sudanic languages and mythological systems. After a 3-to-5000-year pause in SW Asia three Borean language superfamilies emerged, Borean-N (Dené-Caucasian), Borean-M (Eurasiatic) and Borean-R, the latter including language families of SW Asia and Europe as well as SE Asia and Sahul. Alternative short-chronology hypotheses for language evolution, dating of *sapiens sapiens* out-of-Africa and a 'southern route fast track' from SW Asia to Sahul do not appear supported by either mtDNA genetics or archaeology. A hypothesis aligning all language families to the mtDNA phylotree yields a more differentiated and different chronology to the dyadic out-of-Africa dispersion model proposed in Fleming, Zegura, Harrod, Bengtson & Keita (2013).

Introduction.

In linguistics, a macrofamily is a hypothetical phylogenetic grouping of language families based on relationships established by lexicostatistic, multilateral ('mass') comparison or other methods (e.g., Gell-Mann, Peiros & Starostin, 2009). Proponents argue that long-range prehistoric language reconstruction must necessarily begin by applying taxonomic methods before traditional methods of proto-language reconstruction are applied. Examples of proposed macrofamilies include the superfamily Proto-Sapiens (Ruhlen, 1994; Bengtson & Ruhlen, 1994; Trombetti, 1905); models for an out-of-Africa superfamily Borean (Fleming, 2002; 1991; Fleming, Zegura et al., 2013; Starostin, 2006); and smaller groupings, such as Afroasiatic (Greenberg, 1955; 'Afrasian', Fleming, 2002), Nostratic (Bomhard, 2008; Dolgopolsky, 2008, 1998; Starostin, 1999;

Illich-Svitych, 1971), Eurasiatic (Greenberg, 2002; 2000), Dené–Caucasian (Bengtson, 2008; 1997; Nikolayev, 1991); Sino-Caucasian (Starostin, 2004-05a; 2004-05b), Macro-Altaic (Miller, 1991; 1971); Dené-Daic (Starostin cited in Van Driem, 2008), Austric (Bengtson, 2010; Benedict, 1976; van Driem, 2008) and Amerind (Greenberg & Ruhlen, 2007; Greenberg, 1960).

Objections to acceptance of a particular macrofamily may include lack of documentation or scholarship on the constituent languages or disagreements over taxonomic methods or findings. Short-range reconstructionists may object to the general macrofamily approach due to aversion to long-range 'mass' comparisons or belief that an estimated time depth is too great for reconstruction, arguing a limit around 25,000 years ago for glottochronology using a traditional method of Swadesh list binary comparisons (Fleming, Zegura et al., 2013:162), or asserting that the multilateral comparative method has an inherent limit of 6-10,000 years.

With respect to the early dating of language, there is extensive evidence in the archaeological record of symbolic behavior for archaic *Homo sapiens*, Neanderthals and *Homo sapiens sapiens*, and less extensive evidence of even earlier symbolic behavior for *Homo erectus*. As for reconstructing *Homo sapiens sapiens* language evolution, Atkinson (2010) reviews glottochronology methods and observes that the standard method for glottochronology developed by Swadesh places an upper limit on language classification at around 8,000 years and a modified method (Pagel, Atkinson and Meade, 2007; Atkinson, 2010) may extend the limit to 50,000 years or so. If so, such methods appear to be incapable of dealing with predictions for out-of-Africa languages. Atkinson (2011) finds that phonemic diversity supports a serial founder effect out-of-Africa with a cline by distance similar to that of mtDNA, but provides no dating for language family evolution. On the other hand, Perreault and Mathew (2012) find that '*proto-Sapiens sapiens*' language arose in tandem with the emergence of the species. Based on phonemic diversity they calculate that the language of *sapiens sapiens* emerged between 163 and 242 ka, a date range corresponding to the earliest fossil attributed to our species, Omo Kibish, ~195 ka.

One of the obstacles to long-range macrofamily phylogenetics is the absence of a method to establish a comprehensive global chronological sequence for *sapiens sapiens* language evolution. Fleming, Zegura et al. (2013:162) observe "the major problem with linguistic genetic taxonomy is dating or time depth." In this study I aim to attenuate this obstacle. I propose to derive a model for a global language family phylotree drawing on mtDNA archaeogenetics. My literature review indicates some attempts have been made to detect mtDNA correlations for particular geographic areas, e.g. Europe (Soares, Achilli et al., 2010) and Central and South America (Wang, Lewis et al., 2005; Hunley, Cabana et al., 2007), but as yet there has been no global attempt.

Given the vastness of the Y-DNA and mtDNA archaeogenetic literature, I have chosen to focus on mtDNA, conducting a comprehensive review of mtDNA studies of haplogroup frequencies for particular population or tribal samples and currently spoken language. I use the standardized global mtDNA phylotree (van Oven and Kayser, 2009, Build 15, 30 Sep 2012). I use Soares, Ermini et al. (2009) and Behar, van Oven et al. (2012) with their comprehensive reexaminations of mtDNA phylogenetics and calculations of haplogroup TMRCA dates. Archaeology and fossil mtDNA studies are used as an additional check on correlations. My method for deriving correlations for the *sapiens sapiens* language phylotree is primarily one of meta-pattern analysis across disciplines.

There are at least two models for the earliest language families that emerged with *Homo sapiens sapiens* out-of-Africa. Fleming (2002; 1991; Fleming, Zegura et al., 2013) has proposed the existence of a super-phylum, Borean, including ten different language groups, with Afrasian and Amerind as western and eastern anchor groups.

- Afrasian (Greenberg's Afroasiatic) (Semitic, Egyptian, Berber, Chadic, Cushitic, Omotic)

- Kartvelian
- Dravidian
- Sumerian, Elamitic, and a few other extinct languages of the Near East

- Eurasiatic (Greenberg's grouping including Etruscan, Indo-European, Uralic, and Altaic-Mongolian-Tungusic, Japanese-Korean-Ainu, Gilyak/Nivkh, Chukotian, Eskimo-Aleut)

- Vasco-Caucasic (Basque of Iberia, Caucasic of the Caucasus)
- Burushaski and Yeniseian
- Tibeto-Burman (Sino-Tibetan)
- Na-Dené

- Amerind

This list may be grouped into two macrofamilies with the first two clusters combined as 'Nostratic' (Pedersen, 1931, Illich-Svitych, 1971, Dolgopolsky, 1998, Bomhard, 2008) and the third, Dené–Caucasian (Bengtson, 2008; 1997; Nikolayev, 1991).

Gell-Mann, Peiros and Starostin (2009) hypothesizes a 'Borean' that included Nostratic (combining the first two major groupings above) and a second super-family Dene-Daic, subdivided into Dené-Caucasian, the third grouping above, plus Austric, which thus would add a fourth cluster to Fleming's list, Austric, with four clades:

- Austroasiatic (Mon-Khmer)
- Hmong-Mien (Miao-Yao)
- Austronesian
- Tai-Kadai (Daic).

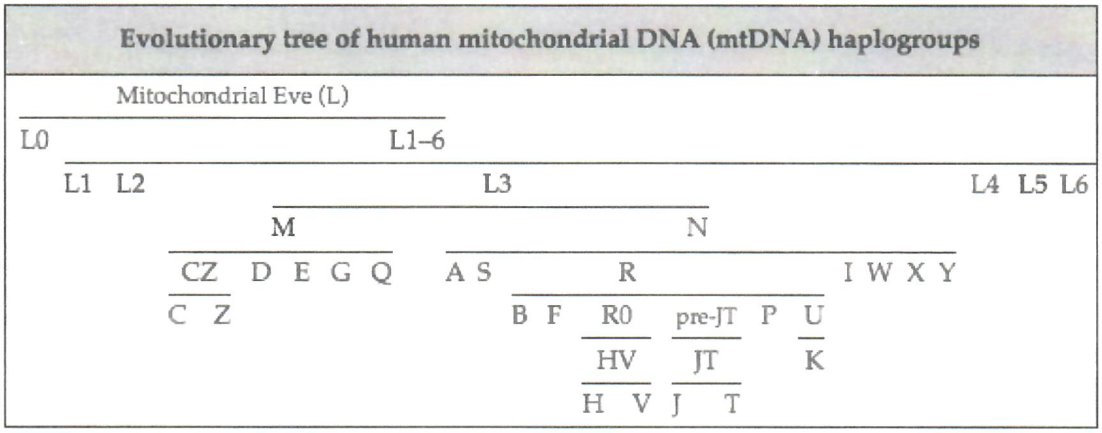
Compare hypothetical tree with Nostratic (Eurasiatic and Afroasiatic) and a supercluster termed 'Borean' (Dene-Caucasian and Austric) (Starostin, 2006) http://starling.rinet.ru/images/globet.png (accessed 15 August 2014). These four clusters cover the major language families outside of Africa, with the exception of a fifth group consisting of Sunda-Sahul languages:

- Pama-Nyungan ('Australian')
- Gunwinyguan ('Australian')

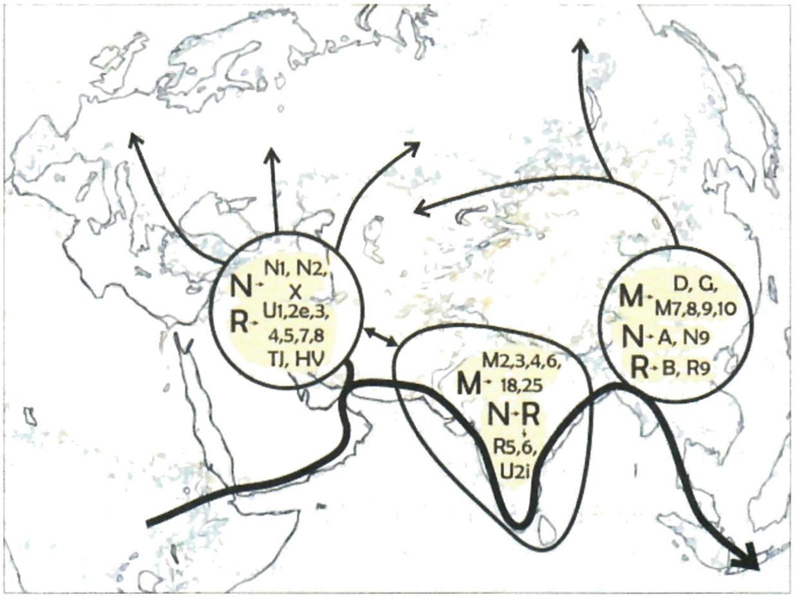
- Trans-New-Guinea
- Papuan, and perhaps
- Paleo-Sundic, including Kusunda of Nepal and Andamanese

In this study, I suggest how all five language family groups appear to correlate to distinctive mtDNA haplogroups and I infer the relative dating of the emergence of these language families based on archaeogenetic TMRCAs (Time to Most Recent Common Ancestor).

A summary of the current basic mtDNA phylotree (after *en.Wikipedia*) follows:


Evolutionary tree of human mitochondrial DNA (mtDNA) haplogroups
Mitochondrial Eve (L)
L0
L1–6
L1
L2
L3
L4
L5
L6
M
N
CZ
D
E
G
Q
A
S
R
I
W
X
Y
C
Z
B
F
R0
pre-JT
P
U
HV
JT
K
H
V
J
T

A generally accepted map of mtDNA homelands and migratory routes out-of-Africa and across Eurasia (*en.Wikipedia,* citing Metspalu, Kivisild et al., 2004) follows.


N· N1, N2, X
R· U1,2e,3, 4,5,7,8 TJ, HV
M· M2,3,4,6, 18,25
N·R
R5,6, U2i
M· D, G, M7,8,9,10
N· A, N9
R· B, R9

The map appears to rely on a short-chronology model of out-of-Africa by a fast 'southern route'. In the study that follows I present a series of maps based on mtDNA haplogroup TMRCAs that suggest the timing of stages for both in-Africa and out-of-Africa homelands and migrations, including the effects of geographic bottlenecks such as the Sinai Crossing, Transcaucasus Crossing to Europe, Zagros Crossing to India and the Ganges Crossing into SE Asia, on both haplogroup migration routes and the emergence of language families.

Hypothesis.

At first glance some language families appear strongly correlated to a basic mtDNA haplogroup, such as African click languages and L0-mtDNA or Eurasiatic and M-mtDNA. Given such prima facie correspondences, a hypothesis that there are no correlations between languages and genetics or mere randomness appears unsupported. Given such obvious correlations, I hypothesized that by and large each basic haplogroup might well correspond to a language family. Further, I hypothesized that the mtDNA phylotree as a whole would correlate to a global African and out-of-Africa language phylotree.

It is critical to note that when I use the prefix 'Pre-' before a language name, it does not designate a reconstructed protolanguage. I used this prefix to designate (a) that a specific mtDNA haplogroup emergence has occurred indicating a population has begun genetic separation and, presumably, associated geographic separation from other populations sufficient to provide a basis for the emergence of a distinct language family and (b) that in current populations this haplogroup appears differentially associated with a particular language family. My method is top-down; 'Pre-' does not mean a 'proto' as in a protolanguage reconstruction. I am not arguing that a haplogroup emergence date is the date for the emergence of a protolanguage. I am suggesting that a particular haplogroup TMRCA date corresponds to the emergence of an ancestral mtDNA lineage that is strongly associated with one particular language family. An associated protolanguage might have emerged around the time of the haplogroup's most recent common ancestor date (TMRCA) or at a later time. Hypotheses about the timing of a language family emergence with respect to TMRCA must be checked on a case-by-case basis. In some cases, especially around 25 ka ('ka' = thousand years ago) or later, particular haplogroups may be associated with a dozen or more different languages, and I leave these aside under phrases such as 'adopted various languages'. These cases may have resulted from the original language family being completely or partially lost.

To take an example of how I use the term 'Pre-', consider my term 'Pre-Basque'. While HV-mtDNA and its descendent H are frequent in Europe, the population with the highest frequency of these haplogroups in Europe occurs in Basque country, depending on the study sample up to 81%. The HV-mtDNA phylotree sequence, TMRCA dates in parenthesis, is: root haplogroup R branched R0 in northern SW Asia (~40 ka). R0 branched R0 subclades across SW

Asia, especially the Arabian peninsula (~13 ka) and branched HV in the NW Fertile Crescent (~25 ka). HV dispersed westward through the Caucasus (high frequencies of HV in NE-Caucasic-speaking Dargin and Avar, 43% and 20% respectively) into Belarus-Ukraine, where it branched HV4 (~14 ka), and along the way H (~15 ka). In addition to the highest frequency of HV and H in Europe, the Basque population also has around 10%U5, which is a marker associated with Uralic languages. This suggests an extended pause in the Belarus-Ukraine area. The migration continued across southern Europe to the Franco-Cantabrian ice age refugium yielding HV4a1 (~10 ka). Genetic isolation of HV4a1a around 5.4 ka may signal the emergence of the current distinct Basque language and culture (Gómez-Carballa, Olivieri et al, 2012; Behar, Harmant et al, 2012; Yunusbayev, Metspalu et al, 2012; Abu-Amero, Larruga et al, 2008). The R0/HV diffusion route with its extended pause in Belarus-Ukraine would predict Uralic borrowings in Basque. Not being a linguist of these languages I leave this prediction aside. I mention this case to stress how I am using the prefix 'Pre-'. When I correlate R0/HV-mtDNA under the label 'Pre-Basque', with a rough beginning date of 50,000 years ago, I am referring to the mtDNA lineage that is differentially that of current Basque language speakers. I leave open the question of protolanguage reconstruction and the question whether, if Basque belongs to the Dené-Caucasian family, the migrating R0/HV-bearing population maintained or adopted a linguistic predecessor of Basque when they crossed the Caucasus bottleneck.

One might speculate when a language emergence occurred subsequent to a haplogroup emergence, but I suggest it probably occurred before the next major haplogroup branching event. For instance, R-mtDNA branches off R2'JT in SW Asia with a TMRCA around 54 ka and subsequently JT-mtDNA, which I suggest correlates to Pre-Semitic-Egyptian, around 47 ka, and then T branches off at around 25 ka. This is roughly a 25,000-year span. When during these 25,000 years Semitic-Egyptian or Semitic and Egyptian emerged I leave as an open question. I suggest that in general language family emergences probably occurred closer to the TMRCA than later branching events. Whether or not one might some day discover a general rule for an average delay time across all major mtDNA haplogroup TMRCAs—say, 5,000 years after—I would argue that the phylotree that I have detected would still maintain its ordered pattern of branches.

The general linguists view on language continuity is that for any given geographic area, languages may have been submerged or completely lost, or abandoned in favor of another language. Worldwide evidence for such is evident in historic records. One cannot assume present-day languages continue ancestral languages over tens of thousand years or even a thousand years. The results of this study challenge this assumption. Just as highly differentiated mtDNA haplogroups are recent, so some languages are recent. However, such a claim is not universal. In the case of African click-languages, an opposite inference for long-range continuity is supported. According to mtDNA genetics and archaeology, Hadza, Sandawe and Khoisan languages differentiated externally from all other language families around 160,000 years ago. East African click peoples split from Southern African click speakers around 100,000 years ago.

Hadza and Sandawe appear to have genetically split by 70,000 (based on L0f lack in Hadza). The full Khoisan tool-and-art package is archaeologically evident in southern Africa by 45,000 years ago. Southern Khoisan (SAK) split from Northern Khoisan prior to 25,000 years ago. We do not know what languages may have been in southern Africa prior to the Khoisan arrival and thereafter lost. This lack of knowledge on our part does not refute genetic and archeological evidence that click language family continuity extends over at least 100,000 years. To date no alternative hypothesis for the apparent relationship between Khoisan and East African click languages has been proposed. Thus, the continuity of any given current language may be short-range or long-range; it must be established on a case-by-case basis and tested against mtDNA and Y-DNA phylotrees and archaeology.

With respect to TMRCA dating different studies sometimes suggest divergent dates for a TMRCA. This may result in linguists eschewing any attempt to use mtDNA TMRCAs. There is also the problem of selecting a molecular clock mutation rate and its dependence on the human-chimpanzee split date ('CHCLA'), which over the decades has been steadily revised upward, so that TMRCAs must be revised to earlier and earlier dates (Harrod, 2013). Whether or not a TMRCA date is secure or 'absolute' is not as important in this study as the relative chronological sequence of the dates for haplogroup emergence. In this regard, Soares, Ermini et al. (2009) and Behar, van Oven et al. (2012), provide dates for haplogroups covering more or less the whole mtDNA phylotree and represent the two most reliable and useful data sets. The former study calculates dates in relation to the Cambridge Reference Sequence (CRS); the latter, from an African perspective. The relative sequence of dates in each study is roughly similar, and I reference the calculated dates for both in my database (Supplementary File, Table 1: mtDNA Database for Archaeolinguistics, online, https://originsnet.academia.edu/JamesHarrod).

I am not a linguist and I am not proposing a standard linguistic proto-language reconstruction. I am suggesting meta-pattern correlations between current population genetic haplogroup frequencies and current languages and crosschecking these against archaeological data and fossil mtDNA studies. This is basically a thought experiment. I present a hypothesis, search for patterns, propose correlations and unfold their logic. I view the overall pattern as a touchstone for linguists to use in ascertaining the sequence by which *sapiens sapiens* languages emerged. This paper revised some of the datings and genetics to language correlations I very tentatively proposed in Harrod (2013). Being neither a geneticist nor a linguist, I anticipate that some genetics-language family correlations may be refuted by proto-language reconstructions. With this caveat in mind, I welcome counterarguments with respect to specific correlations. If it has been said that conflating language and culture and genetics is the cardinal sin of anthropology, then *felix culpa*.

Method.

First, I reviewed mtDNA archaeogenetics and language family literature relevant to *Homo sapiens sapiens* both in Africa and out-of-Africa and generated from this a Master Database table (Supplementary File, Table 1: mtDNA Database for Archaeogeneticlinguistics, online, https://originsnet.academia.edu/JamesHarrod). To date this database table is 181 pages long and contains at rough count 424 ethnic and population mtDNA haplogroup frequency samples and their current spoken language, and 82 fossil mtDNA studies, select archaeological sites and their references. For mtDNA phylotree and haplogroups I used van Oven and Kayser (2009) Build 15 (30 Sep 2012). For mtDNA TMRCA (Time to Most Recent Common Ancestor) dates I relied primarily on Soares, Ermini et al. (2009), and crosschecking dates in Behar, van Oven et al. (2012). Where they do not provide a TMRCA for a particular haplogroup I have cited other studies. In addition, some relevant archaeology is incorporated into the table. As far as I am aware, Supplementary File, Table 1 is the most comprehensive inventory of mtDNA population samples, fossil mtDNA and archaeological parallels available online. It covers the full 200,000 years of *sapiens sapiens* evolution, all macrolanguage families, and many subfamilies.

For each basic mtDNA haplogroup (Hg) I asked what, if any, language family distinctively (as opposed to some other language family) corresponds to it. Generally, current ethnic or population samples are diverse; they have multiple mtDNA haplogroups. To tentatively hypothesize a correlation between an mtDNA haplogroup and language family, I employ several criteria.

Procedure 1. Phylotree. I order the entire master database (Supplementary File, Table 1, http://www.originsnet.org/publications.html) by the standard mtDNA Phylotree and TMRCA for each haplogroup.

Procedure 2. Haplogroup Samples. Under a specific haplogroup, I generally list several population or tribe haplogroup frequency samples with focus on those having highest frequencies for that haplogroup and try to rank order them by frequencies. For a given population or tribe I may list more than one study of that groups mtDNA haplogroup frequencies.

Procedure 3. Hg Population Assignments. In order to construct the master database, I assigned a population sample to a specific haplogroup, e.g., L1 or M or Q3. The rationale for assignment is strongest, of course, if a whole population shares one basic haplogroup, but this is rare. A second criterion for selection is to choose the Hg with the highest frequency in a sample. A third is to select the Hg with the oldest TMRCA. I may use one or both of the latter two criteria depending on contextual factors. In listing a sample in the database I generally place in bold the Hg with the highest frequency in the sample.

Procedure 4. Homeland. I rely on Hg homelands proposed in Soares, Ermini et al. (2009) or other archaeogenetic studies.

Procedure 5. Language Family Correlate. For every population or tribal mtDNA frequency sample I attempt to identify the current language and language family (grey highlight in the

Master Database). Reviewing the currently spoken languages allocated to a particular haplogroup may show a more or less clearly distinctive language family for all the peoples listed under it. In some cases if current language families are diverse, an assignment of the Hg to a language family may be supported or rejected by the logic of mtDNA Phylotree in which it is embedded, circumstantial evidence or known history. In other cases I do not hypothesize any definite correlation to one language family. This phenomenon becomes more frequent as the timeline of the mtDNA phylotree progresses.

If for populations listed under a particular haplogroup, there is a consistent current language family but it is a mismatch for the mtDNA, I assume that the current language is adopted and replaces an earlier language family. One of the most evident mismatches is the M-mtDNA derived D-mtDNA population (Han). By the logic of the phylotree Han peoples should have a Eurasiatic language but their current Chinese language belongs to the Dené-Caucasian macrofamily otherwise correlated to N-mtDNA.

Procedure 6. Archaeological and Fossil mtDNA Context. Archaeological site dates as well as fossil mtDNA studies are noted under particular haplogroups. They contribute to supporting or ruling out particular mtDNA and language correlations and homelands.

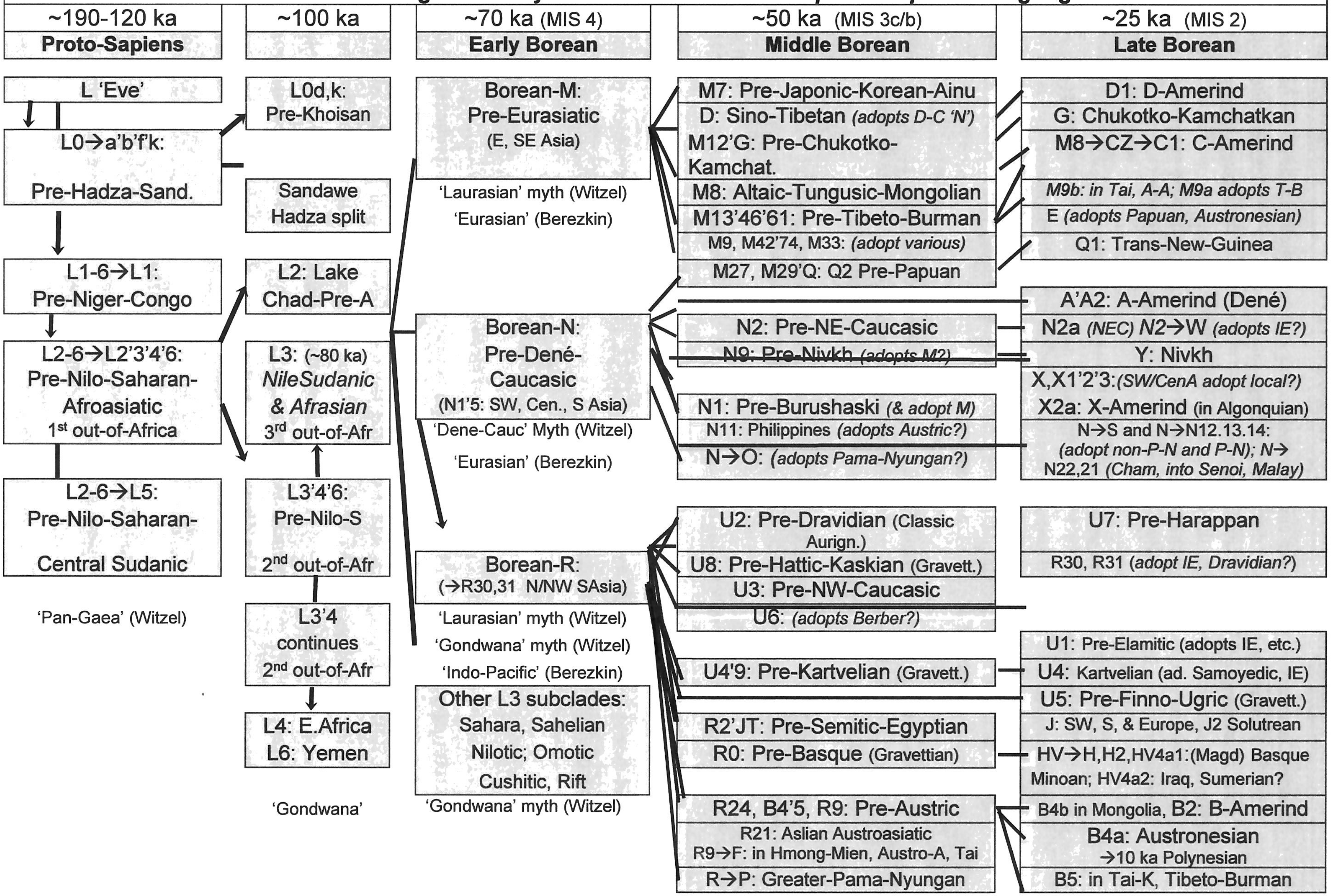
Procedure 7. Mythostratigraphy. Once the language phylotree and language family emergence dates are coordinated, identify mythological systems that may correspond to the phylotree.

Clearly, the correlations I propose are tentative. I reiterate that I propose only a hypothesis, a thought experiment applying the mtDNA phylotree to all *sapiens sapiens* languages. I am not proving or linguistically reconstructing protolanguages or emergence dates. I am developing a meta-model based on TMRCA dates along the mtDNA phylotree against which prehistoric linguistic reconstructions may be guided and tested.

Results.

Table 1 summarizes the meta-pattern-analysis of *Homo sapiens sapiens* language evolution phylotree crossmapped onto the mtDNA phylotree. It is derived from the comprehensive Master Database (Supplementary File, Table 1, online, https://originsnet.academia.edu/JamesHarrod). All of the identified mtDNA haplogroups and correlated language families will be discussed in this results section in chronological order beginning with the emergence of *sapiens sapiens* around 200,000 years ago. Table 1 also suggests very tentative mtDNA correlations to the global evolution of *Homo sapiens sapiens* mythological systems as proposed by Witzel (2011) and Berezkin (2010a, 2010b), and these also will be discussed in greater detail below.

Table 1: mtDNA Archaeogenetic Phylotree Correlated with *Sapiens Sapiens* Language Families



~190-120 ka	~100 ka	~70 ka (MIS 4)	~50 ka (MIS 3c/b)	~25 ka (MIS 2)
Proto-Sapiens		**Early Borean**	**Middle Borean**	**Late Borean**
L 'Eve' L0→a'b'f'k: Pre-Hadza-Sand.	L0d,k: Pre-Khoisan Sandawe Hadza split	Borean-M: Pre-Eurasiatic (E, SE Asia) 'Laurasian' myth (Witzel) 'Eurasian' (Berezkin)	M7: Pre-Japonic-Korean-Ainu D: Sino-Tibetan *(adopts D-C 'N')* M12'G: Pre-Chukotko-Kamchat. M8: Altaic-Tungusic-Mongolian M13'46'61: Pre-Tibeto-Burman M9, M42'74, M33: *(adopt various)* M27, M29'Q: Q2 Pre-Papuan	D1: D-Amerind G: Chukotko-Kamchatkan M8→CZ→C1: C-Amerind *M9b: in Tai, A-A; M9a adopts T-B* E *(adopts Papuan, Austronesian)* Q1: Trans-New-Guinea
L1-6→L1: Pre-Niger-Congo L2-6→L2'3'4'6: Pre-Nilo-Saharan-Afroasiatic 1st out-of-Africa	L2: Lake Chad-Pre-A L3: (~80 ka) *NileSudanic & Afrasian* 3rd out-of-Afr	Borean-N: Pre-Dené-Caucasic (N1'5: SW, Cen., S Asia) 'Dene-Cauc' Myth (Witzel) 'Eurasian' (Berezkin)	N2: Pre-NE-Caucasic N9: Pre-Nivkh *(adopts M2)* N1: Pre-Burushaski *(& adopt M)* N11: Philippines *(adopts Austric?)* N→O: *(adopts Pama-Nyungan?)*	A'A2: A-Amerind (Dené) N2a *(NEC) N2→W (adopts IE?)* Y: Nivkh X,X1'2'3:*(SW/CenA adopt local?)* X2a: X-Amerind (in Algonquian) N→S and N→N12.13.14: *(adopt non-P-N and P-N); N→* N22,21 *(Cham, into Senoi, Malay)*
L2-6→L5: Pre-Nilo-Saharan-Central Sudanic 'Pan-Gaea' (Witzel)	L3'4'6: Pre-Nilo-S 2nd out-of-Afr L3'4 continues 2nd out-of-Afr L4: E.Africa L6: Yemen 'Gondwana'	Borean-R: (→R30,31 N/NW SAsia) 'Laurasian' myth (Witzel) 'Gondwana' myth (Witzel) 'Indo-Pacific' (Berezkin) Other L3 subclades: Sahara, Sahelian Nilotic; Omotic Cushitic, Rift 'Gondwana' myth (Witzel)	U2: Pre-Dravidian (Classic Aurign.) U8: Pre-Hattic-Kaskian (Gravett.) U3: Pre-NW-Caucasic U6: *(adopts Berber?)* U4'9: Pre-Kartvelian (Gravett.) R2'JT: Pre-Semitic-Egyptian R0: Pre-Basque (Gravettian) R24, B4'5, R9: Pre-Austric R21: Aslian Austroasiatic R9→F: in Hmong-Mien, Austro-A, Tai R→P: Greater-Pama-Nyungan	U7: Pre-Harappan R30, R31 *(adopt IE, Dravidian?)* U1: Pre-Elamitic (adopts IE, etc.) U4: Kartvelian (ad. Samoyedic, IE) U5: Pre-Finno-Ugric (Gravett.) J: SW, S, & Europe, J2 Solutrean HV→H,H2,HV4a1:(Magd) Basque Minoan; HV4a2: Iraq, Sumerian? B4b in Mongolia, B2: B-Amerind B4a: Austronesian →10 ka Polynesian B5: in Tai-K, Tibeto-Burman

Notes: (1) Alphanumerics indicate mtDNA haplogroup (Hg). (2) TMRCA dates based on Soares, Ermini et al. (2009) and Behar, van Oven et al. (2012). (3) Designation 'Pre' indicates ancestral Hg emergence, not a proto-language reconstruction. (4) Columns 1, 2 and 3: Witzel and Berezkin proposed myth-strata.

Derived from the literature review detailed in the Master Database (Supplementary File, Table 1, https://originsnet.academia.edu/JamesHarrod) and its inferred language family correlations, highlights of the 200,000-year prehistory of *Sapiens sapiens* language evolution follow. We first look at mtDNA evolution in Africa. It can be reduced to five basic early stages prior to out-of-Africa stages, namely L, L0, L1, L5, L2 and then L3, and they emerge roughly every twenty thousand years (Table 2).

Table 2. Early Stages of in-Africa mtDNA Phylotree Prior to M/N Out-of-Africa

~190 ka	**L-'Eve'**					
~160 ka		**L0**				
~140 ka			**L1**			
~120 ka				**L5**		
~100 ka					**L2**	
~80 ka						**L3**

L-'Eve': Proto-Sapiens-Sapiens. The mtDNA phylotree begins with L-mtDNA ('Eve'), which has a TMRCA at around 200,000 years ago (Behar, van Oven et al., 2012; Soares, Ermini et al., 2009). This closely matches dating for the first fossil evidence attributed to *Homo sapiens sapiens*,

- Omo Kibish, Ethiopia, KHS Early MSA, Levallois and discoidal cores, tools, 1 cordiform handaxe, exotic white-green unknapped opal silicate stone, sourced 20 km away; similar industry AHS with 9 foliates, 1 handaxe, 2 unknapped opal silicate pieces, and BNS with foliates (Shea, 2008; Shea, Fleagle et al., 2004), KHS early *Homo sapiens sapiens* ~195±5 ka (McDougall, Brown and Fleagle, 2005; Brown and McDougall, 2012); thus at border of MIS 7 (244-190 ka) and MIS 6 (190-130 ka).

With respect to language, this would correlate to 'Proto-*Sapiens Sapiens*' ('proto-Human'). The homeland of L is generally considered to be East Africa. With respect to a trait list of symbolic behaviors (Harrod, 2014), Omo Kibish exhibits the collecting and transporting—I would say 'curation'—of non-local exotic stones with an aesthetic quality. This continues a tradition of collection of exotic objects that begins at least by the Acheulian time period.

With respect to NE/E African archaeology circa 190 ka the earliest MSA emerges contemporaneously with Final Acheulian/Sangoan industries, the latter dating from 300 to 150 ka. Final Acheulian/Sangoan sites, more or less securely dated to around 190 ka include:

- Mieso, east-central Ethiopia, Mieso 7 and 31, bifaces, cleavers, ~212±16 ka (de la Torre, Mora et al., 2014; Benito-Calvo, Barfod et al., 2014);
- Sai Island, Nile River, northern Sudan, Site 8-B-11, Level 5, Sangoan, stone circle with 2 slabs with depressions (for grinding?), <223±19 ka, and Level 6, Sangoan, grinding stones, core axes, red and yellow ochre lumps, some with ground surfaces; 10 cm thick sandstone slab, flaked around perimeter, top pecked flat, 10x5cm depression, surrounded by 7 1cm cupules, several chert pebbles with red/yellow ochre adhering, one with black inclusions, suggests symbolic,

>182±20 ka (Van Peer, Fullagar et al., 2003); {JBH: overall shape possible zoomorphic sculpture?].

Cultural interaction between the Early MSA and Sangoan culture traditions is evident, for example the handaxes and collection of exotic stones at Omo Kibish.

L0: 'Pre-Khoisan' and L1'2'3'4'5'6: 'Pre-Niger-Congo-Kordofanian-Nilo-Saharan-Afroasiatic'. At around 160,000 years ago, L-mtDNA ('Eve') branched into L0-mtDNA, which strongly correlates to the click language family of Sandawe-Hadza and later southern Africa bushmen, and L1'2'3'4'5'6-mtDNA (Figure 1). In my review I have found no current population samples still expressing the L1'2'3'4'5'6 haplogroup, nor L0, but only later subclades of the two branches. Aside: based on current population samples Cruciani, Trombetta, Massaia et al. (2011) places root Y-DNA emergence in central and northwest Africa.

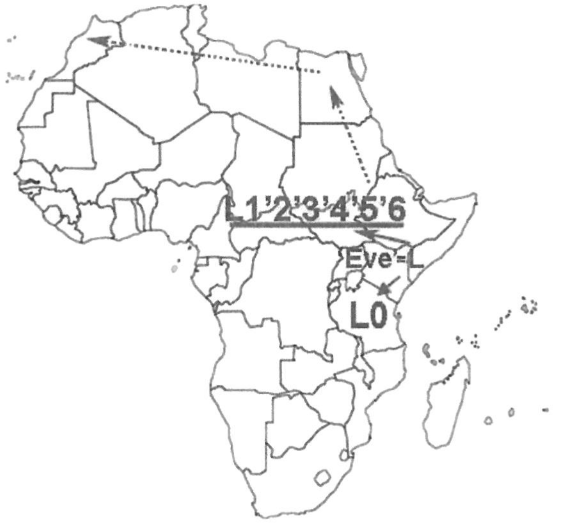


Figure 1: ~160 ka, MIS 6 (190-130 ka), Stage 1 Africa mtDNA Map. Eve'=L originates in East Africa ~195 ka (Omo-Kibish fossil) and branches L1'2'3'4'5'6 in northerly direction and L0 southerly, correlating to the basic split between click languages and all other languages.

Archaeologically, the TMRCA of L1'2'3'4'5'6-mtDNA would correspond to MIS 6 sites; some more or less securely dated include:

- Bir Tarfawi, Western Desert, Egypt, White Lakes, Bed 9, 150-160 ka (Wendorf, Schild et al., 1994) or ~175 ka (Van Peer, 1998)
- Benzú, near Tangier, Morocco, L3b, Mousterian Levallois, 173±10 ka (Ramos, Bernal et al., 2008)
- Ifri n' Ammar, northeast Morocco, Lower O1, MP without tanged tools, 171±12 ka (Richter, Moser et al., 2010)
- Jebel Irhoud, southern Morocco, 4 MNI *Homo sapiens sapiens*, closest morphology though slightly more primitive features to Skhul-Qafzeh, associated with Levallois Mousterian tools, 160±16 ka (Smith, Tafforeau et al., 2007);

These sites suggest that L1'2'3'4'5'6 spread across both central Africa and northern Africa, though this short list of sites does not seem to suggest which area is the homeland and to which it spread.

Final Acheulian/Sangoan industries continue to occur during this time period and presumably interacting with Early MSA peoples. More or less securely dated are:

- Herto, Bouri Formation, Ethiopia, 154±7 to 160±2 ka (Clark, Beyene et al., 2003), Final Acheulian tools, obsidian sourced 289 km away (Negash, Brown and Nash 2011), suggesting complex social trade, exchange, gift-giving (Sahle, Hutchings et al., 2013), 3 MNI early *Homo sapiens sapiens* (McCarthy and Lucas 2014) each with defleshing cutmarks and juvenile bone polishing indicative of mortuary ritual but not cannibalism, and cutmarks similar to those on New Guinea skulls (Clark, Beyene et al., 2003);
- Sai Island, Nile River, northern Sudan, Site 8-B-11, Level 4, Sangoan, >152±10 ka and <182±20 ka, Level 5 (Van Peer, Fullagar et al., 2003);

With respect to language, L1'2'3'4'5'6-mtDNA appears to correspond to the ancestor of all 'non-click' language families. Linguists have grouped languages of the central African corridor, and including some Saharan languages, as a superfamily, variously designated the 'Macro-Sudan belt' (Güldemann, 2011; Clements and Rialland, 2008), 'Niger-Kongo-Nilo-Saharan' (Dimmendaal, 2001; Bender, 2000), 'Niger-Saharan' (Blench, 1995); 'Kongo-Saharan' (Gregersen, 1972), 'Nuclear African area' (Greenberg, 1983, 1963, 1959), or 'Sudansprachen' (Westermann, 1949, 1935, 1911). For the language family correlate of L1'2'3'4'5'6 I suggest the term 'Pre-Niger-Congo-Kordofanian-Nilo-Saharan-Afroasiatic'.

With respect to mythostratigraphy Witzel (2011) proposed a chronological sequence from Pan-Gaea to Gondwana to Laurasian myth strata. To which stratum should one assign the mythologies of click-speakers? If 'L-Eve' is taken as the origin-point for current world mythologies, and it is designated Pan-Gaea, then L0-'click' and L1'2'3'4'5'6-'non-click' mythologies would appear to be the earliest branches of L-mtDNA which might serve for reconstructing Pan-Gaea.

Based on limited evidence from only one site, Herto, one of the earliest traces of *Homo sapiens sapiens* symbolic (proto-religious) behavior appears to be the ritual defleshing and veneration of bones. This may be interpreted as some sort of cult of the ancestors and their bones, a behavior that in recent times has been widespread across central Africa as well as out out-of-Africa, and especially across southern Asia into Sunda-Sahul. Herto appears to continue an Acheulian tradition, where cutmarks on the Bodo Homo erectus skull appears to have been caused by intentional postmortem defleshing (White, 1986).

Archaeogenetics further indicates that southern African click speakers ('Khoisan') diverged from the East African click speakers ≥ 100 ka, and subsequently the northern and southern Africa Khoisan split occurred between 25 to 43 ka (Schlebusch, Skoglund et al., 2012, compare dates in Tishkoff, Gonder et al., 2007). This correlates with their arrival in southern Africa, as evidenced by findings that San material culture (bow-and-arrow, poisons, tool-and-symbol kit) is fully represented at Border Cave, South Africa by around 44 ka (d'Errico, Backwell et al., 2012). Based on its TMRCA, L0f in Sandawe but not Hadza appears to be a token of the split between Sandawe and Hadza by around 70 ka. Those who argue that human language suddenly emerged 25,000 or even 50,000 years ago must explain how it is that southern African Khoisan and eastern African Hadza-Sandawe both have click languages, which are generally accepted as related, if in fact the two populations separated over 100,000 years ago.

L1: 'Pre-Niger-Congo-Mande-Kordofanian (Katla-Rashad)' and L2'3'4'5'6: 'Pre-Nilo-Saharan-(Kadu)-Afroasiatic family'. Around 140,000 years ago L1-mtDNA and L2'3'4'5'6 split

(Figure 2). L2'3'4'5'6-mtDNA has a TMRCA between 149±33 ka (Soares, Ermini et al., 2009) and 139±10 (Behar, van Oven et al., 2012). L1 has TMRCAs between 140.6±33 ka to 128.5±11.1 ka. Thus, the emergence of the two branches occurred during the later MIS 6 (190-130 ka). Highest frequencies of L1 occur among Niger-Congo speakers, for example, Mbenga western pygmies of the Republic of Congo and Central African Republic, for whom DNA samples range from 74% to 100% L1c, which is the earliest dated subclade of L1 (TMRCAs ranging from ~78 to 102 ka). The highest frequency of L1c among non-pygmies occurs among Nzebi, southwestern Gabon, who are Bantu speakers (47%). Lesser rates of L1c occur among the Gabon Fang (29%) and South Cameroon Bassa (24%), Ewondo (21%) and Bakaka (14%), along with varying frequencies of L1b (TMRCA ranging from ~4 to 30 ka, homeland possibly Central Africa), all currently Bantu-speaking populations. Thus there appears to be a robust association between L1 and the Niger-Congo language family. The L1 and L2'3'4'5'6 split thus appears to correlate to the differentiation of a 'Pre-Niger-Congo-Mande-Kordofanian (Katla-Rashad)' family from a 'Pre-Nilo-Saharan-(Kadu)-Afroasiatic family'.

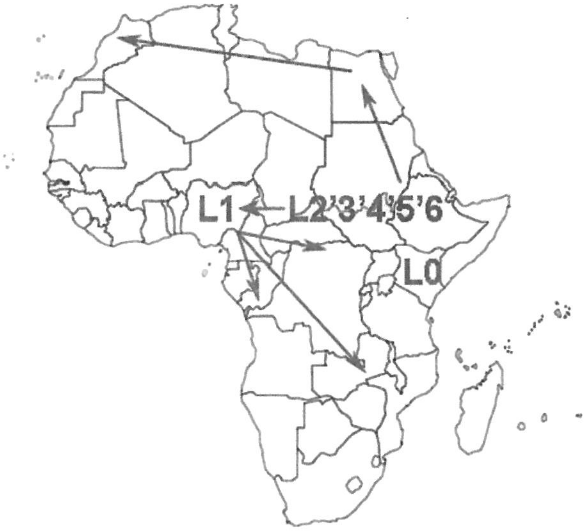


Figure 2: ~140 ka, MIS 6 (190-130 ka), Stage 2 Africa mtDNA Map. L1'2'3'4'5'6 branches off L1 (141x1.1=155 ka), highest frequencies in West Africa, yielding L2'3'4'5'6.

With respect to L1, Niger-Congo speaking groups having only L1c have zero L0, suggesting the exclusion of Niger-Congo origins from 'click' origins. Studies of Mande speaking groups indicate low frequencies of L1c and strong admixture of L2 (L2>L1cb), which may suggest this is a branch of Niger-Congo emerging later than the Atlantic branch. My review found no studies of mtDNA Hg frequency samples for Katla, Rashad or Kadu that might help ascertain their correlations to the mtDNA phylotree.

With respect to archaeology, 'Pre-Niger-Congo' appears to be strongly associated with dates and geographic location for the African Early MSA Lupemban tool-and-art package. Lupemban is considered the regional MSA style of central Africa (Taylor, 2011). On the role of African rainforests in early human dispersals see Mercader (2002). Sangoan Final Acheulian and Lupemban MSA sites occur in Central and West Africa including the Congo Basin, but to date none are securely dated. For the ~140 ka time period, the nearest dated Lupemban sites are:

- Twin Rivers Kopje, Zambia, F-block, Lupemban, choppers, lanceolate bifaces, picks, handaxes, with abundant pigments of multiple colors and a grindstone with incised grooves. ~139 to <178 ka (Barham, 2012; Barham, 2002a; Barham, 2002b; Clark & Brown 2001);
- Kabwe, Broken Hill Mine, Zambia, Lupemban or MSA, bone tools, red ochre, a red-stained 60-mm spheroid (Barham, Pinto Llona et al., 2002; Clark et al., 1947), dated by fauna to 125 ka and AAR on bone to 110 ka (Bada et al., 1974), though based on Lupemban ~130 to 300 ka, (Barham, Pinto Llona et al., 2002), MNI 3 'archaic' *sapiens* fossils, though association to the tool industry is not yet definite (Barham, Pinto Llona et al., 2002; Trinkaus, 2009).

The TMRCA for L2'3'4'5'6-mtDNA around 140 ka corresponds to more or less securely dated early Nubian sites in E/NE Africa such as:

- Sai Island, northern Sudan, Site 8-B-11, Upper Levels 1, 2, 3, Nubian Complex MSA with Lupemban features, <152±10 ka, MIS 5, overlying Sangoan (Van Peer, Fullagar, et al., 2003);
- Taramsa 1, Qena, Upper Egypt, EMP Activity Phase I, Lupemban, lanceolates, foliates, mostly non-Levallois, also Levallois with Nubian ridge, between min. ~117 and max. ~166 ka (Vermeersch, Van Peer & Paulissen, 2010);
- K'one, Ethiopia, Nubian, Centripetal and Bidirectional Levallois, ~140 ka (Rose, Usik, et al., 2011; Kurashina, 1978);

and to the earliest evidence for Aterian in Africa:

- Ifri n' Ammar, northeast Morocco, Upper O1, Aterian MP with tanged tools, earliest evidence for Aterian in Africa, 145±9 ka (Richter, Moser et al., 2010);

and perhaps to the ambiguously dated:

- Gademotta, Ethiopian Rift, Ethiopia, Unit 13, ETH-72-6, Early MSA, obsidian, mostly Levallois reduction, tools mostly foliate/convergent points, denticulates, 4 Nubian Type 1 cores; 1 exotic retouched flake, the largest in the entire assemblage and the only lithic in basalt in this assemblage and all the others from Gademotta and Kulkuletti and suggesting group mobility (Douze, 2012); date between Unit 12, 260±7 ka, and Unit 15, 105±1 ka; perhaps maximum closer to Kulkuletti Unit D Tuff 185±5 ka (Sahle, Hutchings et al., 2013), though I infer that the Nubian Type 1 cores suggest a maximum ~150 ka. I note that the photo of the exotic basalt retouched flake (Douze, 2012: Fig. 70) appears to have features, whether natural or due to flaking, that appear both zoomorphic and anthropomorphic, and if so would make it even more exotic if not symbolic.

With respect to language family, I suggest that these sites would correspond to the 'Pre-Nilo-Saharan-(Kadu)-Afroasiatic'. My review seems to indicate that there are as yet no sites in Southwest Asia around 140 ka that have tool industries evidencing affinities to Africa. On the other hand there is evidence for an early presence of robust *Homo sapiens sapiens*.

- Tabun Cave, Israel, Layer C, Tabun C industry, hominin fossil C2 classed as *Homo sapiens sapiens* with robust features (Quam & Smith 1998;Vandermeersch 1981) though Neanderthal (Trinkaus, 1983, 1984), 135+60/-30 ka (Grün & Stringer 2000), though date must be older (Rink, Schwarcz et al., 2004).

With respect to mythostratigraphy, given my correlation of L1 and Pre-Niger-Congo, I suggest that a search for the earliest Niger-Congo and L1-mtDNA associated myth-rituals might focus on those of Western Pygmies and the Nzebi of Gabon, who have only and high frequencies of L1c, rather than populations with other L1 subclades, such as L1b and L1a1a, which have significantly later TMRCAs and predominate among Bantu and Mande speakers. My master database shows several Bantu-speaking tribes who have L1c and L1b and for whom L1c is greater than the Bantu markers L1a1a+L1b. These include the Fang (Fan), Bassa, Ewondo and Bakaka in the Gabon and South Cameroon area, which is part of the L1 homeland. With the caveat that they have undergone some degree of Holocene Bantu-expansion genetic admixture, their mythologies may still retain survivals of the earliest myth stratum. On the other hand, I would not suggest using any other Bantu tribes in my review database to reconstruct the Middle Paleolithic myth stratum for L1-mtDNA populations. In any event, given the genetic and archaeological correlations discussed above, including the lack of evidence for out-of-Africa diffusions at this stage of *sapiens sapiens* evolution and an L1 homeland in Central and West Africa, I would not suggest using Bantu-expansion mythologies to reconstruct out-of-Africa mythological systems. I leave for future research mythological analysis of Western Pygmy and Nzebi mythology to reconstruct the L1 myth stratum, and only note in passing the former's focus on assuring the benevolence of ancestor spirits and game spirits and lack of a creator or high god.

With respect to mythostratigraphy, my mtDNA and archaeology review does not suggest to me any evidence that might correlate specifically to L2'3'4'5'6 myth structures and so a minimal presumption would be that people associated with this haplogroup may have continued mortuary rituals similar to Herto, and given tool industries with Lupemban features at Sai Island and Taramsa, probably an associated ancestor cult with some similarity to that of Niger-Congo speaking Western Pygmy and Nzebi.

L5: 'Pre-Central Sudanic' and L2'3'4'6: 'Pre-Nilo-Saharan-Afroasiatic (minus Central Sudanic)'. Next around 120,000 years ago, L2'3'4'5'6 bifurcated into L5-mtDNA and L2'3'4'6 (Figure 3). L5 has TMRCAs ranging from ~106 ka to 138 ka; L2'3'4'6, ~111 ka to 115 ka, and thus both lineages emerge during MIS 5e/d (130-106 ka).

L5 appears to have highest frequencies in central Africa, among several Eastern Central-Sudanic speaking populations: Mbuti pygmies (15%) (Tishkoff, 2007 citing Vigilant, 1991) or (10%) (Quintana-Murci, Quach et al., 2008) and Lugbara, West Nile, Uganda, originally from Sudan (14%) (Isabirye, 2010). While the Lugbara sample contains no L0 and may be taken as a proxy for L5, I note that their mtDNA sample shows various admixtures (14%L1*, L1b + 43%L2*, L2a + 21%L3 + 7%L4). Further, Eastern Pygmies, such as Mbuti, Sua and Asoa have zero L1-mtDNA indicating their prehistory is genetically distinct from western pygmies. With respect to Mbuti, a caveat is that the they have 25-30% L0a and thus Tishkoff, Reed et al., (2009) observe that the Mbuti share genetic ancestry with Khoisan peoples, suggesting they lost their initial click language and adopted Central Sudanic. Though linguists such as Ehret (2011) and Blench (2014; Drake, Blench et al., 2011, Bender, 1997) generally class Central Sudanic as Nilo-Saharan and reconstruct it as later than other Nilo-Saharan branches, my review suggests that 'Pre-Central-Sudanic' emerged first, prior to the other ancestral Nilo-Saharan languages. (Note. mtDNA samples for Western Sudanic Bongo-Bagirmi speakers in Chad, the Laka and Boulaia, are predominantly L2a and show no L5, the Sara only 5%, and thus they may be viewed as genetically—as well as linguistically—distinct from the Central Sudanic Lugbara and Eastern Pygmies.)

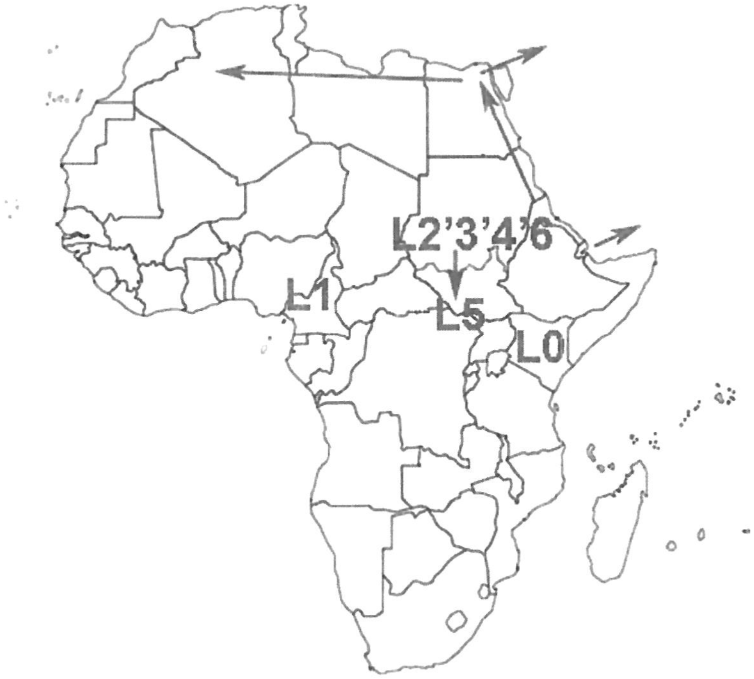


Figure 3: ~120 ka, MIS 5e/d (130-106 ka), Stage 3 Africa mtDNA Map. L2'3'4'5'6 branches off L5 in Central Africa (120x1.1=132 ka) yielding L2'3'4'6 (115x1.1=126 ka). L2'3'4'6 appears to correlate to a first Middle Paleolithic out-of-Africa diffusion, e.g., Abdur Reef, 125±7 ka, and Jebel Faya, U.A.E., ~112 or 123 ka; and Early North African Aterian/Mousterian, Sahara wet-phase, MIS 5e, sites dating 117-125 ka, may have participated in MIS 5e dispersal out-of-Africa as well as material exchanges with Tabun C peoples at Skhul.

As for Central African archaeological sites more or less securely dated to MIS 5e/d that correspond to the L5 TMRCA, my review has only turned up one site, Mumbwa Caves, though given its location the people who left their remains may have had a genetic admixture of L1, L5 and L0.

- Mumbwa Caves, Zambia, Unit VIII: MSA, specularite and red and yellow sandstone for pigments; hearths, windbreak with probably natural 'anthropomorphic' dolomite piece, ~130-170 ka, OIS 5e; Unit VII, MSA with multiple worked pigments (hematite, specularite, limonite, sandstones), 2 heat treated with 'crayon' shape, 107±11 to 130±6.2 ka, MIS5d (Avery, 2003; Barham & Debenham 2000; Barham, Pinto-Llona & Andrews, 2000; Barham, 2000, 1995; Watts, 2009) or Bayesian reanalysis 75-148 ka (Millard, 2008) [JBH: average ~111 ka, MIS 5d].

With respect to mythostratigraphy, I suggest that Lugbara mythology along with that of Eastern Pygmies would appear to be the best bet for retaining survivals of the earliest 'Sudanic' myth stratum as well as retaining aspects of material symbolic behavior evident in the Mumbwa MSA, such as colored pigments and putative natural anthropomorphic stones. I note that the divine power of Lugbara mythology (Middleton, 1979, 1960) is neither a high god or deus otiosus, but a polarity of sacred power, manifesting both as Adroa, a power manifesting in the sky, weather, lightning and associated spirits, and as Adro on the earth, in rivers and streams, a being whose body is cut down the center and who has only one eye, ear, arm and leg and terrible to see. As odd as this may seem to those of a monotheistic or dualistic bent, based on Harrod (2010) I suggest that this Lugbara theology retains strong features of the Late Acheulian symbolic meme, and conversely supports the proposed interpretation of that meme. Like the Western Pygmies, the Eastern have neither a high god nor a deus otiosus, but unlike them have no rituals for ancestors; they do have anthropomorphic game spirits who live in the forest, which is the source of life and where the dead dwell (Sawada, 2001-03). Both the Western Pygmy and Lugbara mythologies appear to place strong emphasis on ritual engagement with life-giving and life-animating forces.

As for L2'3'4'6, the logic of the mtDNA phylotree and language correlations suggest it would associate to a 'Pre-Nilo-Saharan-Afroasiatic (minus Central Sudanic)' language family.
NE/E African archaeology sites during Stage 3 (MIS 5e/d, 130-106 ka), which might correlate to ancestral populations bearing the L2'3'4'6, include multiple regionally distinctive tool industries including Aterian, Nubian, and 'Levallois-Mousterian'. Across North Africa the 'Aterian *sensu lato*' tool industry most frequently occurs during Sahara wet phases, MIS 5a 75-85 ka, MIS 5c (Brørup) 98-110 ka, and MIS 5e 117-125 ka, and less frequently before or after these phases. Aterian sites more or less securely dating to MIS 5e/d (130-106 ka) include

- Dar-es-Soltan II, Morocco, layer 7, Aterian, with 'enigmatic heap of sandstone slabs 1m in diameter and 30 cm high', 121 ka, MIS 5e (Bouzouggar, Barton, et al., 2012; McBrearty and Brooks 2000; Debénath, 1994)
- Dar-es-Soltan I, Unit G1, Morocco, MSA, ~126-130 ka, MIS 5e (Barton, Bouzouggar et al., 2009);
- Bir Tarfawi, southwestern Egypt, Aterian, wet-phase MIS 5e and also later 5c/a (Szabo, Haynes and Maxwell, 1995).

Early North African 'Aterian' peoples appear to have engaged in material exchanges with Tabun C industry peoples at Skhul during MIS 5e/d, and possibly there were actual out-of-Africa migration via the Sinai during this stage.

- Skhul, Israel, Layer B, MP Tabun C industry, *Nassarius gibbosulus* shells perforated in similar manner to Aterian beads (Vanhaeren, d'Errico et al., 2006), pigments selected for yellow, orange, and red hues, 3 pieces intentionally heated to change color from yellow-orange to red (d'Errico, Salomon et al., 2010); 10 MNI H. *sapien sapiens* in shallow burials, S5 with wild boar mandible (Belfer-Cohen & Hovers, 1992; Bar-Yosef & Vandermeersch, 1993); (ESR-Useries) 100-130 ka (Grün, Stringer et al., 2005); (TL) 119±18 ka (Mercier & Valladas, 1994).

Oued Djebbana, Bir-el-Ater, Algeria, the Aterian type site (so far only 14C date >40 ka) evidences perforated shell beads, which were recently reanalyzed; they have the same perforation pattern as Skhul beads, inferring trade exchange or common ancestry across North Africa into SW Asia (Vanhaeren, d'Errico et al., 2006). Further, Hublin and McPherron (2012) and Hublin and Klein (2011) noting similarities in *sapiens sapiens* craniodental fossils at Skhul, Qafzeh and ~40 ka sites Peştera cu Oase, Romania, and Nazlet Khater, Egypt, argue that northern Africa must also be considered as a possible source for the modern human expansion.

Paleontology combined with the Oued Djebbana and Skhul shell bead production similarities imply a genetic and cultural exchange continuum around the Mediterranean across North Africa, the Levant and even into Eastern Europe.

In addition to Aterian, NE/E African Nubian and 'Levallois-Mousterian' MSA sites during Stage 3 (MIS 5e/d, 130-106 ka) would also correlate to ancestral populations bearing the L2'3'4'6. Sites more or less securely dated include:

- Sodmein Cave, Quseir, Eastern Desert, Egypt, Early Nubian MP, 118±8 ka (Mercier, Valladas et al., 1999; Van Peer, 1998);
- Nazlet Khater, Lower Nile, Upper Egypt, Nubian Complex Mid-MP, ~110 ka (Van Peer, 1998);

- Kharga Oasis, Western Desert, Egypt, Mata'na Site G, Bulaq Wadi 3 Locus 1 and MSA Workshop sites, with Levallois-Khargan and Nubian industries, ~110-129 ka (Smith, Hawkins, et al., 2007; Smith, Giegengack, et al., 2004);
- Bir Tarfawi, southwestern desert, Egypt, BT-A Grey Phase W1, ~125 ka, and Grey Phase W2, ~115 ka, arguably Nubian (Van Peer, 1998), or Levallois Mousterian (Schild, comment on Van Peer, 1998);
- Taramsa 1, Qena, Upper Egypt, EMP Activity Phase II, Nubian and Levallois, min.~ 89 and max.117 ka (Vermeersch, Van Peer & Paulissen, 2010);
- Mumba Shelter, Tanzania, lowest level, Early MSA, with bifacial tools and possible red ochre, ~110-130 ka (McBrearty & Brooks, 2000; Mehlman, 1987);
- Abdur Reef, Eritrea, MSA with handaxes at ~125±7 ka (Walter, Buffler, et al., 2000);
- Jebel Faya, U.A.E., MSA industry with small handaxes and bifacial foliates at (OIS5e; OSL mean of 3 dates ~112 or, eliminating 2 outliers, ~123 ka, which is derived from E/NE African façonnage technique (Marks 2009; Armitage, Jasim, et al., 2011);
- Har Karkom, Negev, Israel, HK190a, HK60, HK61, HK83c, HK23 and HK204, with hutfloors, bifacial tools (handaxes) and Levallois flints, designated 'Mousterian of Acheulian Tradition'; HK60 with a large circle constructed by alignment of stones, including four small stone piles (Anati, 2006); no dates; HK190a: possible 'triangular female figurine' (identified and interpreted James Harrod, during Emmanuel Anati Har Karkom Expedition, April 1996);

In addition to material exchanges between North African Aterian and the Levant (Skhul) peoples, the former bearing presumably bearing L2'3'4'6, Abdur Reef and Jebel Faya suggest a second MSA industry with handaxes, presumably bearing a similar haplogroup, participated in the first wave dispersing *Homo sapiens sapiens* out-of-Africa around 120,000 years ago. If the Har Karkom sites were to be dated, they may fall in the same time period as Abdur Reef and Jebel Faya, and if so, provide further evidence and confirmation for this MSA-with-handaxe diffusion out of Africa. While the Aterian material, if not genetic, exchange apparently diffused around the Mediterranean coast and/or Sinai, the MSA-with-handaxe peoples may have crossed over the Sinai, around the coast of the Red Sea and/or over the Bab el Mandeb.

While the Levantine Skhul hominins were traditionally thought to be a dispersal 'dead end', this seems dubious if dates for Homo sapiens sapiens 'with robust features' in South Asia and East Asia are accepted.

- Sihawal I, Middle Son Valley, India, dark-red stratum, Levallois MP (Shipton, Clarkson, et al, 2013) (OSL) 113±9 ka (Haslam, Roberts, et al, 2011); and Patpara I and II, Unit 4 red-brown clay, Levallois MP, rare bifaces (Shipton, Clarkson, et al, 2013; Blumenschine, Brandt and Clark, 1983; reanalyzed Jones and Pal, 2009) (below Toba Ash ~74 ka)
- Zhirendong, Hejiang, Guangxi, South China, no tools, *H.sapiens sapiens*, with robust features (Th/U-series), flowstone over hominin, min. ~106±7 ka (Liu, Jin, et al, 2010; Jin, Pan, et al, 2009)
- Huanglong Cave, Yunxi, Hubei, China, retouched flake and bone tools, *H. sapiens sapiens* teeth, with a few archaic features (TIMS and ICP-MS) min. ~81±1 ka to max. ~101±1 ka (Shen, Wu et al 2013; Liu, Wu, et al, 2010)
- Callao Cave, Luzon, Philippines, demonstrates ability to make open ocean crossings, gracile *Homo sapiens* (U-series) min. ~67±1 ka (Mijares, Détroit, et al, 2010). This date seems too early to be associated with *H. sapiens sapiens* with R-mtDNA in SE Asia by ~58 ka or N by ~56 ka;

and therefore may associate to the L2'3'4'6 First Diffusion Out-of-Africa or the L3'4'6 Second Diffusion Out-of-Africa?

Plausibly, if not a multi-regional evolution, the industries and fossil sapiens sapiens at these sites would seem evidence for the First Diffusion Out-of-Africa extending into East and Southeast Asia.

While there is not yet dated evidence for Nubian Complex, out-of-Africa sites during this time period, future research might verify they also participated in the first wave out-of-Africa. Thus, based on paleontology, archaeology and genetic dates, L2'3'4'6 appears to correlate to the first wave of diffusions out-of-Africa associated with at least two tool-and-symbol kits, a North African MSA Aterian and a NE Africa MSA-with-handaxes.

Given the archaeogenetic correlations, with respect to mythostratigraphy I would hypothesize that the North African Aterian and NE African MSA-with-handaxes peoples who dispersed in this first wave out-of-Africa, whether by the Sinai route, Bab route or circum-Red Sea route, bore mythological-ritual systems that would have been either distinctive or variants of each other and that would have combined features that survived among later Nilo-Saharan (non-Central Sudanic) and Afroasiatic peoples. If this First Diffusion out-of-Africa with North African Aterian-like and NE African MSA-with-handaxes industries reached South Asia, East Asia, Southeast Asia and the Sahul as the above archaeological sites suggest (see also Fleming, Zegura, et al, 2013) this supports the argument for out-of-Africa 'Gondwana' mythological systems across southern Asia into Sahul (Witzel 2011), and also suggests that such systems may have had two or more sub-components corresponding to these two Middle Paleolithic tool industries.

On this point I would note, tentatively, some similarity between the Dar-es-Soltan II Aterian stone construction, 'enigmatic heap of sandstone slabs 1m in diameter and 30 cm high', and the, yet to be dated, Har Karkom HK60 MP of Acheulian Tradition site with 'large stone circle constructed by alignment of stones including four small stone piles', and also the later (MIS 5a) stone pile constructed with sophisticated color and spatial symmetries at Aïn El-Guettar, Tunisia, Units 16 and 17, classed as 'Aterian' or 'Mousterian with bifacial foliates and rare tanged points' (Gruet, 1954), described in more detail below. Possibly this type of Middle Paleolithic symbolic behavior, stone arrangements, is the precursor of such practices in Australia and elsewhere.

L2: 'Pre-Lake-Palaeo-Chad-Afroasiatic' and L3'4'6: 'Pre-Nilo-Saharan' or 'Pre-Proto-Nilo-Saharan'. The next major stage of mtDNA evolution occurs around 100,000 years ago during the MIS 5c/b Saharan wet phase (106-85 ka or 106-95 ka based on nomenclature). L2'3'4'6 branched off L2 and L3'4'6. L2 is the most common mtDNA haplogroup in Africa. The L2 TMRCA ranges from ~88 to 104 ka; the L3'4'6, ~71 to 105 ka. (Note. With respect to L3'4'6, if one conservatively multiplies the TMRCA dates by 1.1 to correct for more recent dating of the chimp/human split date, the result is ~78 to 115 ka, and average 97 ka, which places it squarely in MIS 5c/b.)

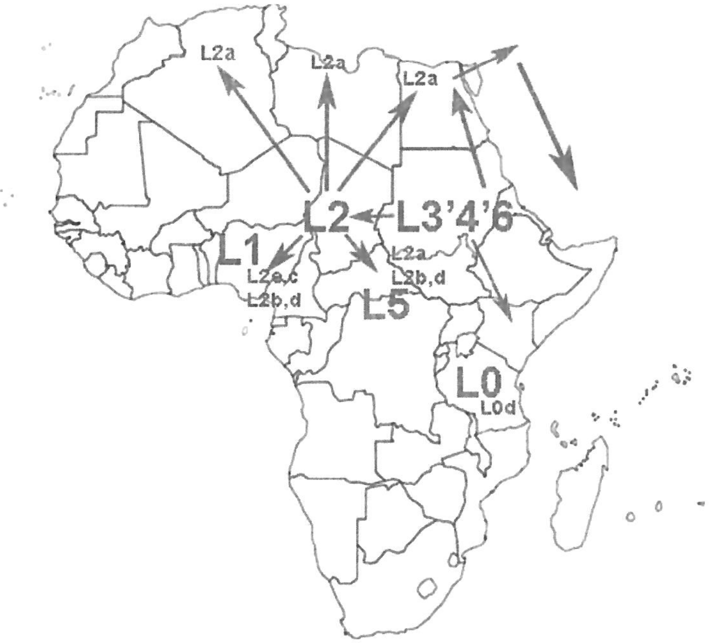


Figure 4: ~100 ka, MIS 5c/b (106-85 ka), Stage 4 Africa mtDNA Map. L0 branches off L0d corresponding to Hadza/Sandawe split and emigration of later southern Africa Khoisan. L2'3'4'6 branches off L2 around Lake Mega-Chad yielding L3'4'6 ~97 ka (averaging Soares-Ermini 105x1.1=115 ka and Behar-van Oven 71x1.1=78 ka), with probable southern Nile origin area. L3'4'6 appears to correlate to a second Middle Paleolithic diffusion out-of-Africa, e.g., Qafzeh, 92±5 ka; Aybut Auwal, Oman, 107 ka, branching off L6 in Arabia.

The L2 subclade L2a'b'c'd emerged ~84 ka (or x1.1 ~92 ka). Its homeland appears to be West or Central Africa (Tishkoff, Gonder et al., 2007). Subsequently, and listing TMRCAs by chronological order, L2a (~79 ka) has high frequencies across North Africa and in Central Africa, e.g., Mbuti, Central Africa (65%); Tuaregs, Mali (27%) and Tunisian Berbers (14%). For these population samples L2a frequencies strongly exceed L1 and later L3 frequencies. If one deletes from these samples back-migrations from SW Asia or Europe mtDNA haplogroups and considers only L haplogroups, the L2a frequencies for Mali Tuaregs is (59%) and Tunisian Berber and Arab speakers sampled in several studies (63%, 60%, 52%, 47% and 46%). While having L2a frequencies less than L3 but greater than L1, L2a also has moderately high percentages among Bongo-Bagirmi Western Central Sudanic speakers currently residing around Lake Chad, including Laka/Kabba Laka (29%), Sara (26%), Boulala (18%). Also may be mentioned here are Songhai speakers who have L2 (20%) but this frequency is less than their frequency of L1 and L3. L2b'c'd dates ~57 ka. L2b has high frequencies in West Africa and across sub-Saharan Africa, L2c in West Africa, and L2d in West Africa and into East Africa. L2e (~47 ka) shows high frequencies in around Lake Chad: Buduma/Yedina, currently Chadic speakers (17% L2e and 30% L2a, b and c) and lesser among Fulani and Fali, North Cameroon, currently Niger-Congo speakers (3-5% L2e).

Rather mixed Palaeo-Chad populations with mtDNA samples that include L2a and L2b, L2d or L2e with L2 greater than L1 or L3 include Mande-speaking groups (North Samo, Mandinga), Senegambian speakers (Fulbe), Chadic (Buduma/Yedina, west of Lake Chad) and Western Saharan Nilo-Saharan (Kanembu, north of Lake Chad). Other mixed L2 subclade groups with L2 less than L3 include Hausa and Kanuri, southwest of Lake Chad. Groups with no L2a but only other southerly originating L2 subclades include the Dogon (67% L2b and c), who speak a language of controversial classification, perhaps an 'early diverging' Niger-Congo ('Mande-like', tonal but no noun class system) and Niger-Congo speaking Tcheboua Fulani, North Cameroon (21% L2b, c and e).

A look at the locations of subclade frequency peaks seems to me to indicate a radiation from a Saharan Lake Palaeo-Chad homeland (Table 3).

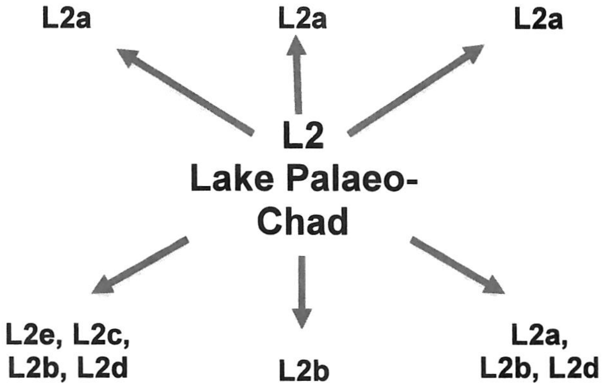


Table 3. Apparent radiation of subclades from L2 homeland, Lake Palaeo-Chad.

OSL dates for Lake Palaeo-Chad range from 97.7±6.5 (Fezzan Basin) to 125±12 ka (Bama Ridge, NE Nigeria), including 119±10 ka (Fezzan Basin), 114±14 ka (Bama Ridge), 108±9 ka (Fezzan Basin), and 107±8 ka (Fezzan Basin) (Drake, Blench et al., 2011). L2 appears to correlate with what might be termed 'Pre-Lake-Palaeo-Chad-Afroasiatic' (L2a), with later populations admixed with L2b'c'd and adopting Niger-Congo or Central Sudanic languages from south of the Sahara.

As noted earlier, the North African 'Aterian' industry *sensu lato* most frequently seems to occur during Sahara wet phases, MIS 5a 75-85 ka, MIS 5c (Brørup) 98-110 ka, and MIS 5e 117-125 ka, and less frequently >50 ka and <150 ka. Based on archaeology and genetics dating, L2 would seem to correlate to an Aterian industry during MIS 5c and L2a to the Aterian during MIS 5a (75-85 ka). Aterian sites more or less securely dated to MIS 5c/b (110-85 ka) include

- Dar-es-Soltan I, Unit G2, Morocco, Aterian, with two ivory objects, one point-like and a plaquette, ~106-119 ka, MIS 5c (Barton, Bouzouggar et al., 2009);
- El Mnasra, Témara, Morocco, Layers 11 and 4base-7, Aterian, with fossil *Homo sapiens sapiens*, bone tools, hearths, worked hematite, *Nassarius* shells, many perforated beads, comparable to Taforalt and other North African and Levantine beads, 109±3 ka [MIS 5c]; Level 4 Upper, Aterian, ~95±9 ka [MIS 5b] (Jacobs, Roberts et al., 2012);
- Grotte des Contrebandiers, Témara, Aterian, *Nassarius* shells, perforated beads (d'Errico, Vanhaeren et al., 2009), ~96-107 ka, mean 103 ka, MIS 5c (Jacobs, Meyer et al., 2011);

Examples of the many North African sites dated by fauna or geology to either MIS 5c or 5a, include:

- Oued Djebbana, Bir-el-Ater, Algeria, the Aterian type site, with perforated shell beads, 14C > 40 ka (Cremaschi, Di Lernia & Garcia, 1998), similar to those at Skhul, 100-135 ka (Vanhaeren, d'Errico et al., 2006);
- Uan Tabu and Uan Afuda, central Acacus, Libya, both Aterian, or Levallois, 65-90 ka (Cremaschi, Di Lernia & Garcia 1998)
- Bir Tarfawi, southwestern Egypt, Aterian, wet-phase MIS 5e and 5c/a (Szabo, Haynes & Maxwell, 1995).
- Aïn El-Guettar, Tunisia, Terrace with spring, classed as 'Final Mousterian with a few tanged points', pile of 60 spheroids with bones, points in spring (Gruet, 1954); (14C) 47±4 and 57±7 ka, i.e., infinite (Aouadi-Abdeljaouad & Belhouchet 2006);
- Nazlet Khater, Upper Egypt, Denticulate Mousterian (K-group), ~100 ka (Van Peer, 1998);
- Bir Sahara, Eastern Desert, Egypt, BS West Lake Phase 2, BS-11 with tanged points, BS-12 and 13 with Levallois and Nubian, with emphasis on denticulates; OSL~104+10/-13, TL range 84±10 to 109±11 (Wendorf, Schild, et al., 1994; Mercier, Valladas, et al., 1999)

These North African industries include one of the most remarkable palaeoart sites at this time period in Africa. Gruet (1954) excavated at the bottom of a spring a pile of 60 spheroids. The El-Guettar construction had 1 tanged point in the base center of the pile, elongated points near the top, at the apex a flint spheroid with white cortex, flaked black one pole and painted with red ochre at its other pole; and two small stone plaques at its base, one a triangle shape the other a lozenge, and rhino and bovid bones scattered throughout the pile. I suggest that the formal symmetrical design of this circular pile of stones may be viewed as an expression parallel to the matrifocal Mali Tuareg creation myth in which a single being, both male and female, descends from above onto an undefined space, and began to move, first towards the East, then counterclockwise creating pairs of same-gendered and opposite gendered beings at each of the four cardinal directions, and finally establishing a world-axis, the male part ascending upward and the female descending downward, with the female initiating and completing the full journey of construction. Berber families and tribes trace their lineages back to ancestors nearly all women and to the ultimate Creatrix (Claudot-Hawad, 2005). This Tuareg creation myth with its high goddess/god who creates the world-cosmos and its axis using a set of complementary opposites and symmetries stands in strong contrast to the apparently earliest central African mythological world with no high god or deus otiosus and focus on ancestor and game spirits conceived in terms of an all-animating life force. I tentatively suggest that this Tuareg creation myth is a mythological survival from 100,000 years ago, and further that in being so similar to the El-Guettar stone arrangement construction, the Tuareg religion corroborates my attribution of L2-mtDNA as a distinctive signal for North African Aterian and related regionalized ethnic groups and the later pre-Afroasiatic language family.

The MIS 5c stage, in which L2 has its TMRCA, is concurrent with the site of Qafzeh, Israel:

- Qafzeh, Israel, Layer XVII-XXIV, Middle Paleolithic Tabun C industry, 18 MNI *Homo sapiens sapiens* fossils, 1 adult bone deposition in niche in limestone wall, Q9 adult female and Q10 6-year old b in 'double burial', Q11 in cavity in soft bedrock with fallow deer antler over hands; 10 Glycymeris marine shells not related to food acquisition, 4 perforated, worked ochre (Bar-Yosef & Vandermeersch, 1993); Q8 burial, red ochre, near broken Levallois core with triangular flat surface used as plaquette incised with mostly parallel stroke marks (Hovers, Vandermeersch & Bar-Yosef, 1997); minimum of 84 ochre pieces at every level, 6 worked, specific hues selected,

> manuported 40 km, associated with burials (Hovers, Ilani et al., 2003; Mayer, Vandermeersch & Bar-Yosef, 2009); TL ~92±5 ka (Mercier & Valladas, 1994, Mercier, Valladas et al., 1995, Valladas, Mercier et al., 1998). Comparison may be made to several other Tabun C industry sites with red ochre, incised stone flake, possible fallow deer burial in red ochre at Nahr Ibrahim, etc.

I note the parallel role of the triangular shape as a design element in the symbolic behavior at both the Qafzeh and El-Guettar sites. As I discussed earlier the Skhul site with date range between 100 to 130 ka, if the lower dating holds it would be dated similar to Qafzeh, and the specific style of its perforated beads would indicate an exchange network between Northern African Aterian and Tabun C cultures. If so this would represent one of two or more Second Diffusions 'Out-Of-Africa Before Out-Of-Africa' of North African Aterian/Mousterian *sensu lato*. This could be called 'Second Diffusion-A'. On the other hand there is not yet any well-dated archaeological remains attributable to the Aterian in SW Asia.

The splitting off of L2 yielded L3'4'6, which has a TMRCA around 105 ka (Soares, Ermini, et al., 2009 or ~97 ka, averaging 105x1.1=115 ka and Behar, van Oven, et al., 2012, 71x1.1=78 ka). While L3'4'6 appears lost in virtually all the current Nilo-Saharan (and Afroasiatic) population samples I could find in my review (see my Master Database), there is possibly one exception to this, the Anuak of western Ethiopia, a Luo-speaking Nilotic (or sometimes referred to as 'Pre-Nilotic') group (Anuak<Luo<Western Nilotic<Nilotic<Eastern Sudanic<Nilo-Saharan). The Anuak are farmers and herdsmen, unlike other Nilotics who are cattle pastoralists. They have mostly managed to continue practicing their traditional religion. Plaster (2011) contains an Anuak sample (n=108), and specifies not the haplogroup frequencies but all their haplotypes. I proceeded to manually determine the haplogroup for each of these individuals using James Lick's mthap Version 16.0 (http://dna.jameslick.com/mthap), and then I calculated the haplogroup frequencies for the entire sample. Interestingly the result indicates that the Anuak—and here I calculate the sample frequencies for L haplogroups only—have more L2a,c,e (24%) than L3 (15%) and L2 frequency equal to L4 (24%), with lesser amounts of L0, L5 and L6. Using the mthap converter, one individual had coding that could be read as L2a or as L2a'b'c'd, and another as either L3x or L3'4'6. Granting that I am a novice at converting haplotypes to haplogroups, and relied on the Lick mthap converter, it appears to me that the Anuak sample potentially shows two very ancient haplogroup survivals. From the fact that the Plaster n=108 sample seems to show both (a) Anuak L2 exceeding L3, which is the reverse of typical southern Nilotic peoples, and (b) possible L3'4'6, I infer that the Anuak mtDNA sample reflects a genetic survival of the original divergence of L2 and L3'4'6.

In a recent study of Ethiopian population groups Pagani et al. conducted a sample (n=23) of Anuak and found that on a neighbor-joining tree analysis Anuak were closest to Gumuz, both on a branch distinct from a branch leading to Afroasiatic speaking Ari, Oromo, Amhara, Tigrai and Somali (Pagani, Kivisild et al., 2012: Fig. S2). Thus the neighbor-joining tree in Pagani et al. (2012) and the Lick mthap converter results seem to corroborate each other. Thus I suggest taking the Anuak as descendents of the earliest speakers of Nilo-Saharan and designating the language family correlating to L3'4'6 as 'Pre-Nilo-Saharan' or 'Pre-Proto-Nilo-Saharan'.

> I note that Ehret (2011) infers that the origin area of Nilo-Saharan is the eastern parts of the southern Sahara, equivalent to the northern Middle Nile Basin; and Ehret (2011) and Blench (2014) both consider proto-Koman (with late Koman, Gumuz and Uduk branches) as the earliest

> branch of proto-Nilo-Saharan. Examining reconstructed subsistence lexicons, Ehret further suggests that proto-Koman, along with proto-Central Sudanic, do not feature horticultural terms, and thus originated with hunter-gatherers. While this may be so, the new genetic typing for a Gumuz sample, as I discuss below, seems to imply that Nilo-Saharan Saharan and Sahelian languages are older than Gumuz.

Archaeologically, during MIS 5c/b stage of evolution, Nubian tool industries occurs in NE Africa and out-of-Africa, with more or less securely dated sites:

- Aybut Auwal, Dhofar, southern Oman, MSA with Nubian Type 1 cores, industry resembling Late Nubian Complex of NE Africa, ~106 ka (Rose, Usik et al., 2011).
- JSM-1, Jubbah paleolake, Nefud Desert, Northern Arabia, Levallois recurrent centripetal, bifacial pieces, some similarity to Dungal and Dineigil Oases, Western Desert, Egypt, latter with Nubian Type 1 cores (Scerri, Groucutt, et al., 2014); Unit-B date 96±9 ka, but stratigraphy ambiguities (Petraglia, Alsharekh et al., 2012; Petraglia, Alsharekh et al., 2011).
- Katotati, Rajasthan, NE Thar Desert, Level S8, MP prepared cores, 1 Levallois-Nubian, retouched tools, bifaces (OSL) 95.6±13.1 ka (MIS 5c) (Blinkhorn, Achyuthan, Petraglia and Ditchfield 2013)

This would presumably represent a diffusion associated with L3'4'6-mtDNA, and I designate it Second Diffusion-B (Nubian Complex) 'Out-Of-Africa Before Out-Of-Africa'. Whether this dispersal migrated to Arabia via the Bab el Mandeb or the Sinai, or most parsimoniously simply dispersed around the Red Sea coast into SW Asia and South Asia may be decided by future archaeological research.

In sum, the divergence of L2 and L3'4'6 around 100,000 years ago and associated archaeology, genetics and languages appears to correlate to the beginning divergence of some features of the Afroasiatic and Nilo-Saharan language families. It at least marks the emergence of 'Pre-Proto-Afroasiatic' and 'Pre-Nilo-Saharan', if not the emergence of proto-Afroasiatic and proto-Nilo-Saharan *sensu stricto*. African language experts such as Ehret (2011) and Blench (Drake, Blench, et al., 2011) argue that the Afroasiatic and Nilo-Saharan language families emerged around 15,000 years ago and possibly correlate to one or another Holocene Epipaleolithic hunter-gatherer industry. The mtDNA Phylotree and its archaeological, linguistic and mythological correlates call into question such a low date for these two proto-languages, since it looks highly likely that aspects of proto-Afroasiatic and proto-Nilo-Saharan emerged by 100,000 years ago or at least by 80,000 years ago—when L2a separated from the other southern L2 subclades.

With respect to my hypothesis of an emergent divergence of 'Pre-Lake-Palaeo-Chad-Afroasiatic (L2)' and parallel branch (L3'4'6) 'Pre-Proto-Nilo-Saharan' 80-100,000 years ago, the following comment based on mtDNA genetic distance analyses may be relevant.

> A phylogenetic tree of genetic distances from inferred ancestral clusters indicates that within Africa, the Pygmy and SAK associated ancestral clusters (AACs) form a clade, as do the Hadza and Sandawe AACs and the Nilo-Saharan and Chadic AACs, reflecting their ancient common ancestries" (4). "Afroasiatic Chadic–speaking populations from northern Cameroon cluster close to the Nilo-Saharan–speaking populations from Chad, rather than with East African Afroasiatic speakers, consistent with a language replacement among the Chadic populations (Tishkoff, Reed, Friedlaender, Ehret, Ranciaro et al., 2009).

As noted earlier, L0 branched L0d around 100,000 years ago, found in Sandawe (5%) but not in Hadza who retain only L0a, and which is predominant in Southern African Khoisan (SAK), such as !Kung (96% L0d+4%L0k). Emergence of L0d appears to correspond to the initial split between East African and Southern African click speakers, with the latter beginning their move to southern Africa from East Africa around 100 ka (Rito et al., 2013; Behar et al., 2008) and arriving in southern Africa with a fully developed San material culture by around 45 ka. The beginning of this migration of click speakers with L0d-mtDNA appears to happen at roughly the same time as the L2 split from L3’4’6.

L3: ‘Nile-Sudanic’ (Nubian MSA and ancestral language to Nilo-Saharan Northern-Sudanic Kunama or proto-Koman, and Eastern Sahelian - Nubian language families); L4: ‘Pre-Nilotic’ and L6: ‘Pre-Yemen’. Haplogroup L3’4’6 began branching its subclades around 85,000 years ago (Figure 5) and L3’4 emerged ~83 ka (averaging Soares, Ermini, et al. [2009] 86x1.1=95 ka and Behar, van Oven, et al. [2012] 64x1.1=70 ka). These dates fall within MIS 5a (85-74 ka).

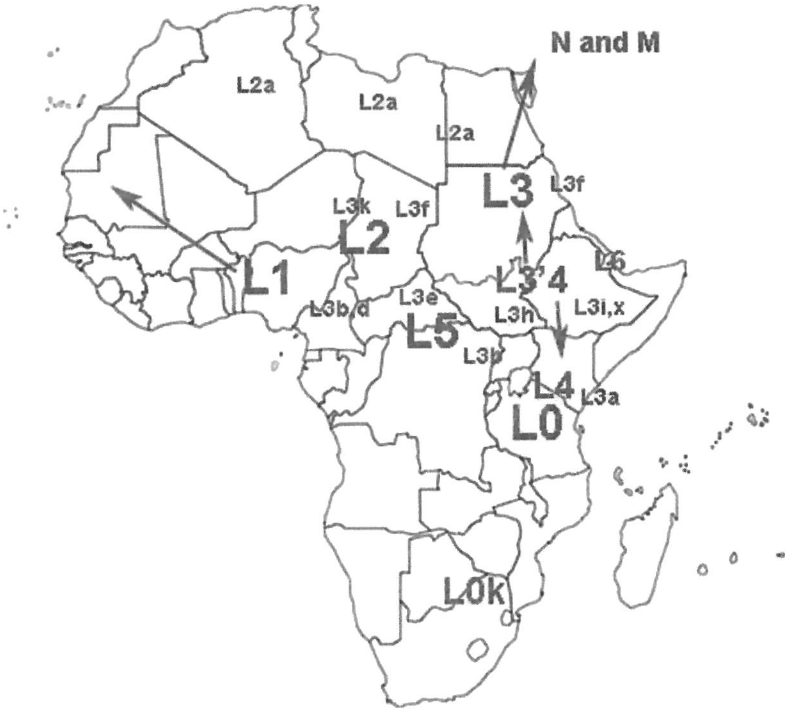


Figure 5: ~80 ka, MIS 5a (85-74 ka), Stage 5 Africa mtDNA Map. Around 83 ka (averaging Soares-Ermini 86x1.1=95 ka and Behar-van Oven L3’4 64x1.1 =70 ka) L3’4’6 branches off L3’4 and around 80 ka L3’4 branches L4, southerly, highest frequencies among pre-Nilotics/Nilotics and Yaaku hunter-gatherers; and L3 (72x1.1=79 ka, a match for 78.3 ka Fu, Mittnik et al., 2013), northerly, highest frequencies in Nilo-Saharan Saharan and Sahelian speakers from the Nile to Lake Chad. L3 branches earliest subclades L3h, ~72 ka (65x1.1=72 ka), highest frequencies among Nilotics and Omotics; out-of-Africa N ~78 (Soares-Ermini 71x1.1=78 ka) and M ~67 (60x1.1=67 ka). Nile Denticulate at Sinai-20 Split Rock, 85±13 ka and again at 62±9 ka; Jebel Qattar. L3 subclades noted at

general locale of highest frequencies (Hirbo thesis 2011). Southern African Khoisan L0k ~40 ka, correlating to San material culture fully represented at Border Cave ~44 ka (d'Errico, Blackwell, Villa et al., 2012).

As for contemporaneous archaeological sites at around 85 ka—and for L3 at 80 ka—this would fall within the MIS 5a (74-85 ka) wet phase. Aterian sites dated to MIS 5a include:

- El Akarit, central coast, Tunisia, Aterian, >80 ka or 90 ka (Reyss, Valladas et al., 2007; Roset & Harbi-Riahi, 2007; Scerri, 2012);
- Dar-es-Soltan I, Morocco, Unit G3, Aterian with foliates, ~68-87 ka, MIS 5a (Barton, Bouzouggar, et al., 2009);
- Grotte des Pigeons, Taforalt, northeast Morocco, Aterian, both with perforated shell beads and other personal ornaments, ~82.5 ka (Bouzouggar, Barton, et al., 2007).
- Ifri n'Ammar, northeast Morocco, Upper OS, 71-95 ka, mean ~83 ka, Aterian with tanged tools, shells, personal ornaments (Richter, Moser, et al., 2010)
- Haua Fteah, Cyrenaica, Libya, lower layers, unique MP industry (Scerri 2012), associated with a fragment of a flute or whistle, initially dated as Eemian by fauna, but OSL ~68-102 ka (Douka, Jacobs, et al., 2013)

Other Mid-MSA industries occur at

- Katanda 9, Upper Semliki Valley, D.R. Congo with bone barbed and unbarbed harpoon points, ~80-90 ka (Brooks, Helgren, et al., 1995; Yellen, Brooks, et al., 1995) or 60-70 ka (Feathers & Migliorini, 2001) [but I note if the highest and lowest outliers of 7 OSL dates is removed, result is ~86 ka];
- Taramsa I-Phase III, Qena, Upper Egypt, Mid-MSA Nubian Complex, Nubian Type 1 cores and points, continuously from ~65 to 84 ka; *H.s.s.* child burial ~69 ka; (Vermeersch, Van Peer & Paulissen, 2010);
- Bir Tarfawi W4, Western Desert, Egypt, Middle Nubian with Nazlet Khater points and Aterian characteristics, ~70 ka (Van Peer, 1998; Vermeersch, Van Peer & Rots, 2005);
- Aduma, Middle Awash, Ethiopia, Levallois, Aduma and Nubian cores, micro-tools, ~80-100 ka (Yellen, Brooks, et al., 2005; Brooks, 2005).

MIS 5a archaeology attests to a continuation of the preceding Second Diffusion 'Out-Of-Africa Before Out-Of-Africa'. Key sites that appear to signal this diffusion include:

- Jebel Qattar 1, Jubbah paleolake, Nefud Desert, northern Saudi Arabia, Unit B, MP Levallois, ~75±5 ka, and Jebel Katefeh 1, Unit H, ~86±11 ka; classed as Tabun-C Levallois (Petraglia, Alsharekh, et al., 2012; Petraglia, Alsharekh, et al., 2011), but the latter with similarities to Dungal and Dineigil Oases, Western Desert, Egypt, latter with Nubian Type 1 cores (Scerri, Groucutt, et al., 2014), which thus appears to continue a flaking industry found at the site of JSM-1, MIS 5b/c. As Tabun-C and North African affinity sites occur at Jubbah in the same timeframe, this suggests possible cultural exchanges.
- Sinai-20 Split Rock, Zarnoq, about 30 km from Taba on the Gulf of Aqaba, 'Nile Denticulate Mousterian', 85±13 ka and again at 62±9 ka (Kobusiewicz, Schild, et al., 2001; Kobusiewicz, 1999; Eddy, Wendorf & Associates1999);
- Gebel Urayf an Naquah, central Sinai, Nubian Type 1 point cores, no date (Schild, 1999; Schild, comment on Van Peer, 1998); located about 24 km from Har Karkom in the Negev Desert with its many MP sites;

- Qadesh Barnea, Wadi El-Qudeirat tributary to El-Arish, NE Sinai, Egypt, MP discoid cores (per se insufficient to identify industry), no date, but by analogy to U-series on mid-MP gravels at Nahal Aqev D35, Nahal Zin, Negev) ~80±10 ka (Schwarcz, Blackwell, et al., 1979); but Goldberg (1986) suggests unconformity warrants rough age ~40-90 ka, given MP/UP transition at Nahal Zin dated ~45 ka;
- Har Karkom, Negev, site HK148b with MP tools including tanged points, and designated the 'Aterian Hut Site', no date (Anati, 2006).
  - The site appears to me to have a floor plan, comparable in dimensions, entrance orientation and artifact arrangements to hut floor plans of Nilo-Saharan Berta, Gumuz, Mao, and other Komuz groups (González-Ruibal, 2006; González-Ruibal, and Fernández Martínez, 2007), and also of Eastern Cushitic Guji-Oromo (van de Loo, 1991).

Considering the above MIS 5a 'out-of-Africa' sites it seems that at least two cultures may have dispersed over the Sinai route, one a Nile 'Denticulate' Mousterian (at some sites classed as 'Nile Mousterian K-group'), and the other an MSA-Nubian Complex tradition, and possibly a third tradition with Aterian features (if HK148b 'Aterian' were dated to this time period), and these interacted with a SW Asian indigenous tradition having a Tabun C industry. In this light, I infer that L3'4 appears to correlate to both cultures diffusing out-of-Africa around the same time, one carrying Nubian industry, the other a Nile Mousterian, and one or both populations also may have diffused an L2-mtDNA subclade (L2a).

With respect to any correlation between L3'4 and language family, my review has not turned up evidence for L3'4 in current population samples, and thus provides no mtDNA evidence for assignation to any particular language family, other than an early Nilo-Saharan or early Afroasiatic. On the other hand, from the evidence for both Nubian Complex and Nile Denticulate Mousterian tool industries out-of-Africa, an archaeogenetic hypothesis would be to associate the Nubian Complex MSA industry with an ancestor of the Nilo-Saharan proto-Northern Sudanic (Kunama) or proto-Koman (Gumuz, Uduk) language families (applying Ehret 2011 terms) and the Nile Denticulate Mousterian with an ancestor of the Afroasiatic Boreafrasian language family. It is tempting to posit proto-Koman at this stage, given the that current Gumuz have predominantly L3 and L4-mtDNA and apparently the highest frequency of these combined (76%) of Nilo-Saharan speaking tribes in NE/E Africa (see Table 4), but lack of L3h in sample makes this questionable (see further discussion of Gumuz below).

Shortly after L3'4 arose ~83 ka, at around 80 ka, L3 and L4 diverged: L3 ~79 ka, using Soares, Ermini, et al. (2009) 72x1.1=79 ka, which then matches 78 ka Fu, Mittnik et al., 2013) and L4 ~79 ka, using Behar, van Oven, et al. (2012), with no multiplier. Later L6 emerged around 20 ka (Behar, van Oven et al., 2012). L6 has peak frequencies among Yemeni (12% and 26% for L's only) and is also frequent among eastern Ethiopians. L6 is said to be recent in Yemeni and have its homeland in East Africa (Soares, Ermini et al., 2009; Kivisild, Reidla, et al., 2004). While at first glance it might reflect L3 out-of-Africa, its late date, restricted typology and lack of N and/or M autochthonous lineages in Southern Arabia does not support a 'southern route' model for out-of-Africa (Abu-Amero, Larruga, et al., 2008).

L4 occurs only in East Africa, which would thus be its homeland (Soares, Ermini et al., 2009; Tishkoff, Gonder et al., 2007). L4 has highest frequencies among Nilotics, such as Nuer (18%) and Acholi (15%), some Omotic speakers, such as Hamer (18%), some Cushitic, including

Yaaku, Kenya, hunter-gatherers (32%) and Tigrai (13%). This suggests the L4 homeland may be more specifically the Southern Sudan/Omo area.

The homeland for L3, the parent of N and M and all subsequent out-of-Africa mtDNA lineages, is either East Africa or NE Africa; to date genetics does not offer a more specific location (Soares, Alshamali et al., 2012). On the other hand, I suggest the differentiation of L3 and L4 appears to correspond to L4 more southerly in East Africa and L3 more northerly along the Nile corridors. The highest frequencies of L3 occur in Nilo-Saharan Saharan and Sahelian speakers from the Nile to Lake Chad. Such a homeland and L3 TMRCA of ~79 ka would corresponds to in-Africa archaeological sites among those listed above under L3'4, such as:

- Taramsa I-Phase III, Qena, Upper Egypt, Mid-MSA Nubian Complex, Nubian Type 1 cores and points, continuously from ~65 to 84 ka; *H.s.s.* child burial ~69 ka; (Vermeersch, Van Peer & Paulissen, 2010);
- Bir Tarfawi W4, Western Desert, Egypt, Middle Nubian with Nazlet Khater points and Aterian characteristics, ~70 ka (Van Peer, 1998; Vermeersch, Van Peer & Rots 2005);

With respect to TMRCAs of L3 subclades there appear to be four clusters chronologically, roughly dating 65, 55, 40 and 25 ka. The earliest subclade cluster includes L3h, N and M, with L3h, ~69 ka (average Soares-Ermini 67 ka and Behar-van-Oven 59 ka = 63x1.1=69 ka). Out-of-Africa N dates about ~75-78 ka (Soares-Ermini 'N in South Asia 71 ka x1.1=78 ka; Soares-Alshamali ~62 ka seems to be a clock violation with their date N→R 'in South Asia' 67 ka, and this would imply N actually emerges between 74 and 79 ka). M dates somewhat later at ~67 ka (Soares-Ermini 61x1.1=67 ka). Highest frequencies of L3h occur among Omotics and Nilotics. Root N and M occur only in SW Asia.

To discern if there were any overall patterns in genetic data and current languages that might yield some inferences about the emergence, homeland or language families associated with L3-mtDNA, I reviewed mtDNA studies of East and Northeast African population groups. I reanalyzed published haplogroup frequencies to focus on L-haplogroups only, deleting back-to-Africa haplogroups of Eurasian origin. I selected samples with highest frequencies of L3 and L3h and I note frequency ratios among L3, L2, L4 and L6. Grouping ethnic groups by language, six clusters became more or less apparent, each with fairly distinct haplogroup frequency patterning (Table 4). Boattini, Castri, et al. (2013), using far more sophisticated principle component analysis and model-based clustering techniques, and focusing on Afroasiatic speakers, appear to me to have independently confirmed four of the six clusters.

Table 4: Combined L3≥30% and/or L3h>9% Frequency *(based on Ls only, subtracting Eurasian)*

	L0	**L1**	**L5**	**L2**	**L4**	**L3**	**L6**	*ratio L3 / L2*
Early Nilo-Saharan: L2>L3, L4>L6 (L3h=~65ka, N and M= ~60ka, ergo **o-o-A?)**								
Anuak *nilotic (White Nile)*	*12%*		*15%*	*24%* 1 L2a'b'c'd?	*24%*	*15%hfexb (1%h)* 1 L3'4'6 or x?	*8%*	*inverse*
Saharan and Sahelian: L3>L2, no or rare L5, L4, L6 (**L3h**=~65ka, N and M= ~60ka, ergo **o-o-A?)** (**L3f**= ~50ka; L3e= ~40ka)								
Kanuri *pro-shr*	*8%*	*3%*		*15%*	*2%*	*66%hfedb (2%h) (20%f, 17%e,d)*		***4x***
Laal *nc+chad*	*18%*	*9%*		*9%*		*63%hfed (9%h) (9%f, 36%e)*		***6x***
Sara *wc sud*	*16%*	*5%*	*5%*	*26%*		*37%hfedb (5%h) (11%f,e,d,b)*		*1.5x*
Nubian *e shl*	*26%*	*7%*	*7%*	*22%*	*1%*	*35%hfi (2%h) (20%f, 16%i)*		*1.5x*
[JBH: This is the Nile Corridor for out-of-Africa as well as E-W Sahel/Sudanian Savanna Corridor.]								
Komuz: L3, L4 but no L1, L2 or L6 (L4= ~80ka; L3f= ~50ka; L3i= ~44ka; L3x= ~35ka; ergo, not o-o-A?)								
Gumuz *komuz*	*12%*		*12%*		*24%*	*52%fxb (no h) (20%f, 24%x, 8%b)*		*infinite*
Semitic, Chadic (Boreafrasian): L3>L2, no L4, L5, L6; no or low L0 (**L3h**=~65ka; N, M=~60ka, ergo **o-o-A?**) (compare Boattini cluster C) (L3f= ~50ka; L3e= ~40ka; **L3x**= ~35ka)								
Egyptian *sm Alexandria*	*10%*	*25%*		*16%*		*50%*hfexdbc (3%h) (2%x, 20%f, 12%e)*		***3x***
Gurna		*6%*		*12%*		*53%*x (no h)*		***4x***
[JBH: This is Northern Africa/Sinai corridor for out-of-Africa.]								
Cushitic, Semitic: L3≥L2, L4>L6 *(**L3h**=~65ka; N, M=~60ka, ergo **o-o-A?**) (compare Boattini cluster A)* (L4= ~80ka; L3a= ~55 ka; **L3f**= ~50ka; **L3i**= ~40ka; L3x= ~35ka; L3d= ~30ka; L3b= ~20ka; **L6**= ~20 ka)								
Agaw *cc*	*16%*	*2%*	*2%*	*26%*	*8%*	*43%hfixdb (4%h) (14%f, 12%i, 6%d)*	*2%*	*1.5x*
Yemeni *sem*	*15%*			*15%*	*6%*	*38%hfiexdb (2%h)*	*26%*	*2.5x*
Afar *ec*	*20%*	*3%*	*2%*	*36%*	*2%*	*34%hafiexdb (8%h)*	*3%*	=
Amhara *sem*	*17%*	*2%*	*5%*	*31%*	*12%*	*31%hfixd (3%h)*	*2%*	=
[JBH: This is region of Blue Nile Corridor for out-of-Africa and/or via Afar and Bab-el-Mandeb.]								
Omotic: L3>L2, L6>L4 (**L3h**=~65ka, N and M= ~60ka, but probably not o-o-A) (compare Boattini cluster B1)								
Maale *n om*	*6%*	*4%*	*6%*	*12%*	*19%*	*49%hafixb (2%h) (3%a, 36%x)*	*4%*	***4x***
Dawro *n om*	*13%*	*1%*		*17%*		*46%hafiexdb (11%h) (4%a, 18%x)*	*5%*	*2.5x*
Hamer *s om*	*20%*			*20%*	*20%*	*30% (30%h)*	*10%*	*1.5x*
[JBH: This is SW ETH, Omo River Basin to Lake Turkana with no corridor for out-of-Africa.]								
L3 Nilotics: L3≥L2, L4>L5, noL6 (**L3h**=~65ka; L3 subclades moving south) (compare Boattini cluster B2)								
Turkana	*24%*		*7%*	*11%*	*13%*	*46%hafiexb (9%h)* (2%a)		***4x***
Samburu	*22%*		*8%*	*14%*	*12%*	*45%hfixdc (9%h)*		***3x***
Luo k	*22%*		*8%*	*14%*	*8%*	*45%hafeb (6%h) (2%a)*		***3x***
Nuer *w nilot*	*10%*		*15%*	*20%*	*20%*	*36%hfeb (15%h) (10%fe)*		*1.5x*
Acholi u	*9%*		*9%*	*27%*	*18%*	*36%heb (9%h) (18%b)*		*1.5x*
Maasai k	*27%*		*12%*	*11%*	*17%*	*33% hafiexb (4%h) (7%ax)*		***3x***
Dinka	*12%*	*6%*	*6%*	*30%*	*14%*	*30%hfie (14%h) (9%f)*		=
Datoga tz	*56%*			*6%*	*13%*	*26%hfb (20%h)*		***4x***
[JBH: Nuer, northern-central Sudan, and Dinka, Southern Sudan: this is the White Nile Corridor.]								
Note. L3h has geographic diverse sub-lineages requiring caution in relying only on HVS-1 (Rosa & Brehm, 2011) Note. Gumuz Hgs (with gratitude, Pagani personal communication Dec 2014 re Pagani et al. 2015 in preparation)								

Given Table 4 and all the preceding L3 tables, I then searched each cluster for a distinctive subclade above and beyond L3h, such as subclade high frequency or feature distinguishing that cluster from other clusters, which would serve as its distinctive marker. I suggest the markers summarized in Table 5.

Table 5. Hypothetical Correlation of L3 Subclades in Current Population samples to Nilo-Saharan and Afroasiatic Language Families with TMRCA-based Emergence Dates

	L3h+**L3x (~35ka)** **EGYPT** (Semitic A-A) (Nile Corridor)	(Note. No L3h in Beja Northern Cushitic A-A) (Red Sea Coastal Corridor)
L3h (~65ka) (L3h homeland?) **OMO** (Omotic A-A) (Omo River Basin)	L3h+**L3f (~50 ka)**+(i)(~40 ka) **NUBIA-SAHARA** (Saharan and Sahelian N-S) (Nile Corridor)	L3h+**L3d (~30 ka)**+(i)(~40 ka) **CUSHITE** (Central Cushitic, Semitic A-A) (Blue Nile Corridor)
	L3h+**L3e(~40 ka)+b(~20 ka)** **NILOTIC** (Eastern, Western and Southern Nilotic N-S) (White Nile Corridor)	L3h+**L3a (~55ka)** **RIFT** (Eastern Cushitic Burunge and Southern Cushitic Yaaku A-A) (Rift Corridor) (Note. No L3h in Oromo, Yaaku Eastern Cushitic, Tigray Semitic or Ongota A-A)

Method: (a) examine the L3h population sample clusters, identify additional subclade(s) that appear to be markers signaling that cluster, e.g., predominant subclade, high frequency of subclade or other feature distinguishing it from the other clusters.

TMRCA dates are based on Soares, Ermini, et al. (2009) and/or Behar, van Oven, et al. (2012). None of these dates are corrected by a 1.1x or 1.2x multiplier suggested by recent studies arguing for redating of the CHCLA.

Afroasiatic and Nilo-Saharan language associated clusters are supported by different methods of cluster analyses in Boattini, Castri, et al. (2013) and phylotree in Pagani, Kivisild et al. (2012).

Every sample in Table 4, except Gurna Egypt has L3f, but L3f's highest frequencies appear to occur in samples from Nilo-Saharan Kanuri and Nubians. Frequencies of L3i are not very dissimilar between Nubians and Cushitic speakers. L3a appears especially distinctive for Eastern Cushitic Afar at 2% versus 0% for other Cushitic and Semitic populations, and for Southern Cushitic Burunge (16%) and originally Eastern Cushitic Yaaku (26%). Raising a question about the main linguistic speculations for Ongota, Ongota genetics suggests they were originally northern 'click-language' speakers, who appear to have later adopted Nilo-Saharan and subsequently some features of Afroasiatic, thus causing considerable exasperation for linguists.

Neither Table 5 nor Table 4 appears to provide any definite clue as to the homeland of L3. On the other hand, if we consider that L3'4 breaks into L3 and L4, the latter in East Africa, this may suggest that L3 has a homeland north of East Africa. Table 5 focusing on L3h suggests that the Omo Basin is either the homeland or remnant of the oldest L3 subclade L3h. If so, I suggest that we may consider that L3M and L3N branched off further north along the Nile Corridor or Sinai Crossing out-of-Africa. The tables appear to provide no strong support for a Bab route out-of-Africa for L3M or L3N.

L3M and L3N out-of-Africa with admixture of L2a: variable combinations of 'Nile-Sudanic' and 'Boreafrasian' languages (Nubian Complex and Nile Denticulate MSA industries). As their TMRCAs are only a few thousand years apart during MIS 5a (L3'4 ~83 ka; L4 ~80 ka; L3 ~79 ka), what I inferred earlier based on archaeology, genetics and language with respect to L3'4 seems to me equally valid for L3-M and N out-of-Africa. There is evidence for both Nubian Complex and Nile Denticulate Mousterian tool industries dispersing out-of-Africa. A likely hypothesis would be that the Nubian Complex MSA industry correlates to the Nilo-Saharan proto-Northern Sudanic (Kunama) or proto-Koman (Gumuz, Uduk) language families, and/or, by the time of L3 ~79 ka, the Proto-Saharan-Sahelian and Proto-Eastern Sahelian (Nubian) languages, applying linguistic terms from the language phylotree in Ehret (2011). To designate the ancestral language family that appears to correlate to the Nubian Complex industry out-of-Africa bearing L3 or its subclades M and N, whether by a Sinai or circum-Red Sea Crossing, I find it necessary to invent a term, and I suggest 'Nile-Sudanic'. The Nile Denticulate Mousterian out-of-Africa may correlate to the Afroasiatic Boreafrasian language family, perhaps with a strong Northern African L2a component. In inferring these two population dispersals out-of-Africa and correlated archaeology and language families, I am not equating L3M and L3N respectively to each of the two archaeological tool traditions. Current mtDNA population genetics shows that Nubian populations and Northern African/Egyptian populations both have high frequencies of L2a and L3 mixture, so I suggest inferring that the populations who made the Nubian Complex and Nile Denticulate Mousterian tool industries already had some degree of such haplogroup diversity and perhaps even language diversity.

With respect to the evolution of mythological structures, I would similarly infer that populations dispersing out-of-Africa, whether by the Sinai route, Bab route or circum-Red Sea route, dispersed with myth-ritual systems that combined to greater or lesser extent components of the two major northern and central African myth-systems, which evolved prior to the TMRCA of L3 and which may be termed 'North African' (correlating to L2, L2a) and 'Sudanic' (correlating to L5, L4) or 'Nile-Sudanic' (correlating to L3 and its subclades). The Sudanic system emphasizes ancestral and game spirits, where 'spirit' means a life-giving, life-animating and life-enhancing forces or energies, with neither a high god nor deus otiosus; the North African, a creative power, female or androgynous, who organizes a cosmos with four directions, complementarities and world-axis and establishes the nature of life's unfolding as one of balance and complementarity of polar or gender opposites. Archaeology and mtDNA genetics suggests that both of these mythological systems dispersed out-of-Africa during MIS 5a around 80 ka.

Again, I do not necessarily associate these two mythological traditions with populations predominantly either M or N-mtDNA, but I do suggest that myth-ritual admixtures may have occurred among these populations, especially as they resided in SW Asia prior to dispersals through the Transcaucasus and Zagros bottlenecks. How and to what extent these two myth-ritual systems might correspond to Gondwana and Laurasian mythological structures (Witzel 2011) is a topic for future research.

M, N and R-mtDNA: Early Borean. The period around 75,000 years ago (early MIS 4 ~74-59 ka) witnessed the divergence of M, N, and R-mtDNA lineages in Southwest Asia. As noted earlier, I date N about ~75-78 ka (Soares-Ermini 'N in South Asia 71x1.1=78 ka; Soares-Alshamali ~62 ka seems to be a clock violation with their date N→R 'in South Asia' 67 ka, and this would imply N actually emerges between 74 and 79 ka). I suggest the date for the branching of R-mtDNA from N at around 74 ka (Soares-Alshamali R 'in South Asia' 67x1.1=74 ka). I date M somewhat later at ~67 ka (Soares-Ermini 61x1.1=67 ka). These dates I propose, around 75 ka, using the x1.1 multiplier produce excellent agreement with the archaeology of NE Africa and Arabia, and are similar to recent conclusions of archaeologists and geneticists (Petraglia, Haslam, Fuller, et al., 2010; Reyes-Centeno, Ghirotto, Détroit, et al., 2014).

These dates suggest that the *sapiens sapiens* dispersal out-of-Africa endured a roughly 5,000 year pause in SW Asia, during which N and M differentiated subclades and N branched off R-mtDNA. Two geographic bottlenecks, the Transcaucasus and Zagros Crossing, extant Neanderthal and other archaic species, the Toba supereruption (~74 ka), as well as the positive subsistence landscape of the 'Persian Gulf Oasis' (Rose, 2010; Rose, 2007) presumably contributed to this delay. A *sapiens sapiens* 'fast track' on some sort of 'southern route', e.g., around 60-50 ka (Mellars, Gori et al., 2013) and earlier studies proposing and African LSA out-of-Africa ~45 ka do not appear supported. Recent genetic studies using various methods argue for such a pause in SW Asia (Xing, Watkins, Hu et al., 2010; Amos & Hoffman, 2009). Such a pause in SW Asia before diffusion north and east through the bottlenecks might be compared to the pause at the later Beringia crossing bottleneck.

Reflecting on Fleming's Borean hypothesis, my meta-review of archaeogenetics, archaeology and language correlations suggests that the three major Borean language families map fairly precisely onto the earliest out-of-Africa mtDNA branches, M, N and R. Eurasiatic languages are strongly associated with M-mtDNA and its branches; Burushaski-Caucasic-Dené languages with N-mtDNA, and Afroasiatic, Kartvelian, Dravidian, Elamitic, etc. with R-mtDNA. Interestingly, Austric and Pama-Nyungan seem most closely associated with R-mtDNA. Noting the controversial hypotheses for a so-called 'southern route' to the Sahul, I suggest that the current mtDNA phylotree appears to me to imply that Austric and Pama-Nyungan descendants are actually on a descendent branch of R, and what I term Borean-R. Here I leave aside for future discussion questions regarding the extent of Denisovan or erectus genetic admixture, technological regime hybridizing, or admixing with languages of archaic *sapiens*.

The earliest branching of subclades of Borean-N, M and R, which occurred around 70 ka to 60 ka, I designate as 'Early Borean'. Using the relative chronological order of TMRCAs in Soares, Ermini, et al. (2009) ['*SE09*'] or Behar, van Oven, et al. (2012) ['*BO12*'], uncorrected for proposed redating of a CHCLA based mutation clock, the 70-60 ka earliest differentiations of Early Borean (each listed in approximate chronological order) include:

Early Borean-N (homeland SW Asia)

- N in South Asia (71.2±16 ka, *SE09*); N → N1'5 (57±5 ka, *BO12*)
- N in W Eurasia (61.9±11 ka, *SE09*)
- N in East Asia (58.2±14 ka, *SE09*)
- N in SE Asia/Sahul. N → N11 (56±4 ka, *BO12*)

Based on these dates N appears to have taken about 15,000 years to disperse over South Asia, Western Eurasia, East Asia and SE Asia/ Sahul, and in that geographic order. No simple 'rapid coastal southern route' is evident here, and, apparently ironically for that hypothesis, dispersal into Western Eurasia occurs 5,000 years prior to SE Asia. I explore more evidence for this and for my hypothesis that N correlates predominantly with the Burushaski-Caucasic-Dené macrofamily when I further discuss N subclades below.

Early Borean-R. N → R (geographical homeland not yet determined, 56.5±2.1 ka, *BO12*; but see earlier TMRCAs below) (Fossil DNA, R*, Ust'Ishim, Irtysh River, Russia, ~43-47 cal ka; see below)

- R in South Asia (66.6±14 ka, *SE09*)
  - R → R31 (64.5±14 ka, *SE09*; 54.9±3.1 ka, *BO12*), with frequency peaks in Rajasthan and Sri Lanka (Karmin, thesis, 2005)
  - R → R30 (64.0±15 ka, *SE09*; 53.6±4.0 ka, *BO12*), with frequency peak in Gujarat, NW to Central India (Karmin, thesis, 2005); R30 → R30b (51.0±4.7 ka, *BO12*), with highest frequency in Vedda, Sri Lanka (Ranaweera, Kaewsutthi et al., 2014)
- R in W Eurasia (59.1±12 ka, *SE09*)
  - R → R2'JT (54.7±12 ka, Near East, *SE09*; 53.7±5.7 ka, *BO12*)
  - R → U (54±11 ka, *SE09*)
- R in SE Asia/Sahul (Australia/Melanesia, 58.4±8.4 ka, *SE09*; or 64.6 ka, Jinam, Hong, et al., 2012)
  - R → P (54.9±3.1 ka, *BO12*), Sahul: Australia, Papua NG, Melanesia
- R in East Asia (54.3±13 ka, *SE09*)
  - R → T16189C! (Caspian-Baikal; as 'R→B' 51±13 ka, *SE09*), highest frequency in Mongol, Southern China, Japan (rare)

Based on these dates R appears to have taken about 13,000 years to disperse over South Asia, Western Eurasia, SE Asia/ Sahul and East Asia, and in that geographic order. Again no simple 'rapid coastal southern route' is evident here. This view is supported by a remarkable recent discovery (Fu, Li, et al., 2014), which provides evidence for the earliest fossil mtDNA out-of-Africa, at Ust'Ishim, Irtysh River, Russia, bearing root-R with a novel SNP not in any other R subclades in current populations, and dating ~43-47 ka. This study notes that in terms of current subclades of root-R, the R haplotype of the fossil is most closely related to mtDNA subclades P, B, F, T and J. When I entered the 12 SNPs in this study into the online James Lick mthap haplogroup assigner, it showed the best matches were root R (11 matches, 1 extra SNP) and next best were R30, P, R2'JT and R(T16189C, which is the parent to B4'5) (also 11 matches, but 1 mismatch, and the 1 extra). Apparently, this individual's lineage with root-R, prior to these subclades evolving, migrated into Western Siberia, and presumably taking one of the likeliest routes to get there, such as from the Persian Gulf Oasis north to the main 'Silk Route' to Central Asia or the 'Silk Route Indus Spur' to Central Asia, and from there migrating further north via the Aral Basin or Irtysh or other tributary to the Ob River. This is further confirmation of my hypothesis herein that the Borean-R language macrofamily had a homeland around the Persian Gulf Oasis and/or NW South Asia and then branched subclades into Western Eurasia, SW Asia, SE Asia/Sahul and Central Asia/East Asia; and a further refutation of the single rapid coastal route to Sahul hypothesis.

Interestingly, the early dispersals of N-mtDNA and R-mtDNA occur in roughly the same west to east sequence over roughly the same period of time. (N disperses in 15,000 years, R in 13,000; N disperses to East Asia, then SE Asia/Sahul, while R the converse; however, given the standard deviation of the TMRCAs these differences seem insignificant.) If this was the case, it raises the question of whether peoples with N and R migrated together? Did the earliest tribal group(s) out-of-Africa have a dual organization of its society that segregated and mixed these two haplogroups, or were they simply two separate diffusions?

I explore more evidence for my hypothesis that R correlates predominantly with speakers of Afroasiatic, Kartvelian, Dravidian, Elamitic, Austric and Pama-Nyungan families when I further discuss R subclades below.

Early Borean-M

- M in East Asia (60.6±13 ka, *SE09*)
  - M → M12'G (57±14 ka East Asia *SE09*; or 47.3±3.6 ka, *BO12)*
- M in South Asia
  - M → M42'74 (~55 ka, India/Australia divergence (Kumar, Ravuri, et al., 2009) or 49.6±8.1 ka, *BO12*)
- M in SE Asia/Sahul (49.4±10 ka, *SE09*; or ~64 kya, Jinam, Hong et al., 2012)

(Note. The wide variance in TMRCA in these studies of the earliest M dispersal reflect variously proposed problems in dating the M lineage compared to N and R lineages, such as tuning the molecular clock, variable mutation rates, possible impacts of the Toba supereruption, and need for more population samples from SE Asia, etc.)

Based on these TMRCAs, the Early M dispersal appears to begin from East Asia, from there to South Asia, and later into SE Asia/Sahul, and taken about 12,000 years. It is possible that the Toba supereruption attenuated an earlier presence of M in South Asia. According to these TMRCAs, M in South Asia, which is associated with current Austro-Asiatic speakers, appears to be a back-migration from SE Asia. Once again, no simple 'rapid coastal southern route' is evident here. I explore more evidence for this and for my hypothesis that M correlates predominantly with speakers of the Eurasiatic macrofamily when I further discuss M subclades below.

Keeping in mind the caveat that the date for the emergence of a particular protolanguage may likely will be later than the mtDNA TMRCA, the genetic and language family correlations I hypothesize as 'Early Borean' at around 70 ka are summarized in Table 6.

During this same stage, circa 70 ka, additional subclades of L3 emerge in Africa, which are associated with subsequent emergences of Sahara-Sahelian, Nilotic, Omotic, Cushitic and Chadic languages families. (For more details for L3-mtDNA subclades, TMRCAs and genetic-language family correlations, see Master Database, Supplementary File, Table 1, https://originsnet.academia.edu/JamesHarrod).

Table 6: ~70 ka (MIS 4) 'Early Borean'			
Early Borean-N	~71 ka (S Asia); ~62 ka (W Eurasia); ~58 ka (E Asia); ~56 ka (SE Asia/Sahul) *Dené-Caucasic (Burushaski)*		
N1'5	~57 ka	(N1: Central Asia, S Asia, SW Asia, Europe; N5: South Asia)	
Early Borean-R	~67 ka (S Asia); ~59 ka (W Eurasia); ~58 ka (SE Asia/Sahul); ~54 ka (E Asia) *Afroasiatic, Dravidian, Hattic, Austric, P-N*		
R31	~65 ka	(adopts Dravidian or I.E.)	
R30	~64 ka	(adopts Dravidian or I.E.)	
Early Borean-M	~61 ka (E Asia); ~55 ka (S Asia); ~49 ka (SE Asia/Sahul) *Eurasiatic*		

With respect to the on-going debates about 'Australoid' groups in southeastern Asia, these groups do not appear to show any similarities in extant mtDNA lineages. Their diversity is characterized by distinct markers among population samples, namely M31 and M32 in the Andamanese, N11b in the Mamanwa of the Philippines, M21a and R21 in the Jehai and Kensiu from West Malaysia, and haplogroups P, Q, S, and O in the Melanesians and Australian Aboriginals (Jinam, Hong, et al., 2012). Thus so-called 'negrito' populations are not a single out-of-Africa dispersal, but multiregional convergences under selection for tropical environments, which is supported by paleontology, fossils, and a variety of DNA studies using different techniques (see e.g., 2013, Special Issue on Revisiting the "Negrito" Hypothesis, *Human Biology* 85,1).

M, N and R-mtDNA: Middle Borean. The By around 50,000 years ago (MIS 3c/b), based on mtDNA phylotree and its TMRCAs, it appears that roughly 28 new mtDNA haplogroups had emerged. I list them below, noting dates, populations with high Hg frequencies, and current languages [in brackets] associated with those populations.

Middle Borean-N [N correlating to Burushaski-Caucasic-Dené macrofamily] (Fossil DNA, 'macro-N' (branch for N1a,b,c, X, I, W) in Gravettian Paglicci12, ~29 cal ka.)

- N in South Asia (71.2±16 ka, *SE09*)
  - N → N1'5 (57±5 ka, *BO12*)
  - N1'5 → N1 (54.2±13 ka, *SE09*; or 52±6 ka, *BO12*) Central Asia, SW Asia, S Asia, Europe [currently mostly I.E. speakers]
    [4.5%N1 and 2.3%N2 in Hunza, Burushaski speakers, but also 25%M]
  - N1'5 → N5 (37±8 ka, *BO12*) India, Madhya Pradesh Sahariya [I.E.]
- N in W Eurasia (61.9±11 ka, *SE09*)
  - N → N2 (44±7 ka, *BO12*) SW Asia, W Eurasia, includes Dargin, Chechen [Northeast Caucasic]
  - N → X (31.8±13 ka, *SE09*) Avar (15%X), Druze (16%X1'3+11%X2), Georgian (8%X)
- N in East Asia (58.2±14 ka, *SE09*)
  - N → N9 (49.1±12 ka, *SE09*; or 45.7±7.9 ka, *BO12*) root SW Asia, then Central Asia to E Asia, Udegey (29%N9+8%N9b), Japan (7%N9a,b)
- N in SE Asia/Sahul
  - N → N11 (56.3±3.6 ka, *BO12*); N11b (6.7±4.6 ka) in Mamanwa, Philippines [Austronesian]
  - N → S (53.5±5.5, *BO12*; or 25.4±5.2, Hudjashov, Kivisild, et al., 2007) only in Australia, Tasmania [Pama-Nyungan and non-Pama-Nyungan]
  - N → O in Australia (48.0 ka, van Holst Pellekaan, Ingman et al., 2008) (Fossil DNA, O1a-mtDNA in aboriginal hair, southern West Australia burial, 100 BP) [adopts Pama-Nyungan]

As observed earlier, N appears to have taken about 14,000 years to disperse over South Asia, Europe, East Asia and SE Asia into Sahul, and in that geographic order. No simple 'rapid coastal southern route' is evident here, and, apparently ironically for that hypothesis, dispersal into Western Eurasia occurs prior to SE Asia.

N1'5 branched N1 at ~54 or ~52 ka. In current population samples it has highest frequencies among the Mazandar, SW Capsian, northern Iran (14.3%); Sindhi, Pakistan (8.7%); Caucasus (8.7%); Lebanon (9%); Finland (8.3%) and Central Asian Kurds (7.3%), populations that currently are mostly I.E. speakers. In addition to 25%M, 4.5%N1 and 2.3%N2 occur in Hunza, who are Burushaski speakers. I suggest correlating N1 to an ancestor of Burushaski. These frequencies appear to reflect a dispersal encountering the Zagros Crossing bottleneck, and taking northerly migrations over the Silk Route Indus Spur to the Central Silk Road, thence west to Southern Caspian, and later splitting (a) back into Mesopotamia and (b) taking the Caucasus Silk Road Spur into Europe and (c) from Indus/Central Silk Road east into Tajikistan, Uzbekistan.

N1'5 branched N5 around 17,000 years later in India, ~37 ka, with highest frequency in the Sahariya, northern Madhya Pradesh (23%). This may reflect a successful Zagros bottleneck crossing into India and taking either a Ganges route across northern India or turning south along the coast and dispersing east over the Narmada River corridor. This hypothesis has archaeological support, as the TMRCA roughly correlates to the recent discovery of India's earliest microblade industry along the Narmada River, Mehtakheri, Madhya Pradesh, Unit 2 ~48 ka, Unit 1 ~35 ka (Mishra, Chauhan & Singhvi, 2013). Maji, Krithika and Vasulu (2008) and Palanichamy et al. (2004) argue that N5 may be autochthonous for India.

N branched N2 around ~44 ka, with high frequencies among Sindhi, Pakistan (17.4%); Kurds in Middle East (10%) and in Asia (8.5%); Finland (9%); Gujarati, India (8.8%); Romania (6.4%); Caucasus (5%); Slovenia (4.8%); Shugnan Tajikistan (4.5%); Pathan, Pakistan (4.5%); Iran (4.1%); Uzbek (2.4%); Hunza (2.3%), many of which populations are currently I.E. speakers. Since I.E. is generally accepted by linguistics as very recent, even Holocene in origin, for the ancestral language family of N2, I suggest taking as a clue the trace N2a (~24 ka) in Daghestan Dargin (1%N2a or 4%N) and Chechens (<1%N2a), who are speakers of Northeast Caucasian. If so, N2 correlates to Northeast Caucasian. Similarly to N1's haplogroup frequency samples, N2 appears to reflect Zagros Crossing bottleneck, taking the Indus Spur to the Central Silk Road and thence (a) west to Southern Caspian, then the Caucasus Spur into Europe and (b) east into Tajikistan and Uzbekistan.

N disperses into East Asia (~58 ka), and appears to branch a Middle Borean subclade N9 (~49 ka), and its later Y branch. Having high frequency among Nivkh, I suggest correlating N9-Y to a 'Pre-Nivkh'. At some point in time, the Nivkh ancestors appear to have abandoned their Borean-N language and adopted a language from the Eurasiatic language family. Early N dispersed into SE Asia (~56 ka), with its N11 branch, the descendents of which seem to have adopted Austric languages. S and O-mtDNA diffused into Sahul, ~53 ka and ~48 ka respectively, where S seems to have adopted non-Pama-Nyungan languages and O adopted Pama-Nyungan languages.

Middle Borean-R [R correlating to speakers of Afroasiatic, Kartvelian, Dravidian, Northwest Caucasic, Austric and Pama-Nyungan families]

- R in South Asia (66.6±14 ka, *SE09*)
  - R → R31 (64.5±14 ka, *SE09*; 54.9±3.1 ka, *BO12*), with frequency peaks in Rajasthan and Sri Lanka (Karmin thesis 2005) [adopt I.E. or Dravidian]
  - R → R30 (64.0±15 ka, *SE09*; 53.6±4.0 ka, *BO12*), with frequency peak in Gujarat, NW to Central India (Karmin, thesis, 2005); generally N/NW South Asia,
    - R30 → R30b (51.0±4.7 ka, *BO12*), with highest frequency in Vedda (39%R30b+R8a1a3), other Sri Lankans (Ranaweera, Kaewsutthi et al., 2014); → R30 → R30a (19.3±6.9 ka, *BO12*) in Central Tharu tribes, Nepal (range: 20% to

3.5%) and R30b1 (14.2±5.3 ka, *BO12*) in Eastern Tharu, Nepal (5%); [adopt I.E. or Dravidian]

    - R → R6 (51.1±16 ka, *SE09*; 43.9±9.6 ka, *BO12*), with frequency peak in Bharia, Madhya Pradesh [Dravidian] (18.4%); Kashmir (5.3%); Tamil Nadu (3.9%); Rajasthan (2.8%); Tharu (2.5%) [adopt I.E. or Dravidian]
    - R → R8 (42±16 ka, *SE09*; 32.8±6.9 ka, *BO12*), in Vedda (see R30b above)
- R in W Eurasia (59.1±12 ka, *SE09*) (Fossil DNA, 'root-R' in fossil Ust-Ishim, Irtysh, ~45 cal ka; R in fossil Fumane, ~41 cal ka.)
    - R2'JT (54.7±12 ka, *SE09*; 53.7±5.7 ka, *BO12*), homeland SW Asia
        - R2'JT → JT (50.3±12 ka, *SE09*; 47.0±6.5 ka, *BO12*) Egypt (27%J+T), Palestine; Iraq;
            - JT → J (32.6±11 ka, *SE09*; 34.3±4.9 ka, *BO12*, but 43.3/34.5 ka, Pala, et al., 2012), SW Asia general (12%), Iran (13%); Europe (11%); Egypt (9%); J in Solutrean Nerja, Málaga fossil DNA, 17-20 ka;
            - JT → T (26.8±9 ka, *SE09*; 25.1±4.7 ka, *BO12*) homeland Caspian?; Swanetia, Georgia [Svan<Kartvelian] (10.4%T+4.2%T1); Egypt (15%); Palestine (13%),; Syria (12%); Iran (10%)
        - R2'JT → R2 (41.0±16 ka (Metspalu et al., 2004); 13.7±6.9 ka, *BO12*) Al-Mahra, East Yemen (12%); Mazandar, SW Caspian [I.E.] (9.5%), Brahui [Dravidian} (7.9%)
    - R → U (54±11 ka, *SE09*) homeland SW Asia (Fossil DNA, U near basal-R, not related to any current subclade of U, in Mal'ta MA-1, child burial, ~24 cal ka.)
        - U → U2 (54±13 ka, *SE09*; 43±4 ka, *BO12*); South Asia autochthonous (Metspalu et al., 2004), Kubachi, Daghestan [NE Caucasian] (24%); Iran (2.4%). (Fossil DNA, U2 in Eastern Aurignacian Kostenki14 burial, ~36-39 cal ka;
            - U2a,b,c(=U2i) (~23, 29, 39 ka, *BO12*), Irula, Tamil Nadu [Dravidian] (50%); Gondi [Dravidian] (36%); Uttar Pradesh Brahmins [I.E.] (27%); Sindhi (17.3%), Pathan (15.9%); Uttar Pradesh (15.3%); Sri Lanka (12%); Pakistan (mixed) (11%); Karnataka (10.6%); Kashmir (10.5%) [adopt I.E. or Dravidian]
        - U → U8 (50.2±11 ka, *SE09*; 43±4 ka, *BO12*); Asia and Europe; U8a (37±14 ka, *SE09*) SW Europe (2.8%); U8b'K (46±11ka, *SE09*) (Asia/Europe). (Fossil DNA, U8 in Gravettian-Pavlovian Dolní Vestonice DV13, ~31 cal ka.)
        - U → U4'9 (~43±12 ka, *SE09*; 37±6 ka, *BO12*), Central Asia, Caucasus, Europe; U4'9→U9 (25.7±6.6 ka, Pakistan, *SE09*) →U9a Andhra Pradesh, Ethiopia, U9b, Pakistan, with possible African origin; and U4'9→U4 (21±10 ka, *SE09*) Kalash [I.E.] (34%), Ket [Dené-Yeniseian] (29%), Nganasan [Samoyedic] (21%), Tundra Nenets [Samoyedic] (13%'U'), Tubalar [Altai] (15-18%), Western Siberia

(Mansi, Nentsi, Nganasan, Ket average) (17%), Mansi [Ob-Ugric] (16%), Pakistan (14%), Volga-Ural peoples (9.7%), Hazara [I.E.] (9%), Swanetia (8.3%) and Georgians [Kartvelian] (8%)

    - U → U3 (~41±12 ka, *SE09*; 33±6 ka, *BO12*), SW Asia; Lur Zagros Iran (18%) [I.E.], Jordan (15%); Adygei (14 or 5%), Abazins (12%), and Kabardin (10%) [latter three NW Caucasian speakers]
    - R → R0 (~39±15 ka, *SE09*; 40±11 ka, *BO12*), SW Asia; Marsh Arabs (0.7% R0 + 6.9%R0a) [descendents of Sumerians]; Persians (2%). (Fossil DNA, 'either R0 or HV' in Gravettian Paglicci25, ~27 cal ka) [Pre-Basque]
- R in SE Asia/Sahul (Australia/Melanesia, 58.4±8.4 ka, *SE09*; or 64.6 ka, Jinam, Hong, et al., 2012)
  - R → P (54.9±3.1 ka, *BO12*), Sahul: Australia, Papua NG, Melanesia
    - P → P4 (53.0±4.4 ka, *BO12*) Australia, Melanesia
      P4b (40.1±7.1 ka, *BO12*) SW and No Australia [Greater-Pama-Nyungan]
    - P → P8 (~39.8 ka, van Holst Pellekaan, Ingman et al., 2006), Yuendumu Warlpiri, central desert, Northern Territory [Pama-Nyungan]
    - P → P1 (32.9±6.2 ka, *BO12*) Melanesia: Irian Jaya highlands [adopts Trans-New-Guinea, Papuan or Austronesian]
  - R → R9 (47±12 ka, *SE09*; or 46.7±7.7 ka, *BO12*) Hmong-Mien (20%)
  - R9 → R9c (46.7±6.3 ka, *BO12*) Philippine Batek Negrito [Austronesian] (58%)
    - R9c → F (43±11 ka, *SE09*; or 42.79±5.6 ka, *BO12*) Temiar Senoi [Aslian<Austroasiatic] (43%F1a1a); Hue Vietnam [Mon-Khmer<Austroasiatic] (29%F*/F1a); Thailand [Tai-Kadai] (20%F) and oddly Ket [Yeniseian] (24%)
  - R9 → R9b (38.5±8.7 ka, *BO12*) Semelai Aboriginal Malay [Aslian Austro-Asiatic] (28%)
  - R → R21 (47±12 ka, *SE09*) Jehai and Kensiu Semang and Temiar Senoi [Aslian<Austro-Asiatic];
- R in East Asia (54.3±13 ka, *SE09*) (Fossil DNA, fossil B* in Tianyuan, China, ~40 cal ka.)
  - R → B4'5 (49.5±6.6 ka, *BO12*) → B4 (44±12 ka, *SE09*), Hmong-Mien; Vietnam [Mon-Khmer]; Polynesia, Micronesia, Madagascar; e.g., Vanatu [Austronesian] (40%)

As observed an above discussion of Early Borean-R, R-mtDNA appears to have taken about 13,000 years to disperse over South Asia (~67 ka, earliest branches R31 ~65 ka and R30 ~64 ka, suggesting a Ganges or Narmada corridor east), Western Eurasia (~59 ka), SE Asia into Sahul (~58 ka), and East Asia (~54 ka) and in that geographic order. Again no simple 'rapid coastal southern route' appears to be evident in these Early Borean R subclades and their dates. Considering the above Middle Borean haplogroups and their TMRCAs around 50,000 years ago, it seems that R-mtDNA has a much more complex pattern of differentiation and dispersal than does N or M-mtDNA. Given the TMRCAs, R emerges in South Asia and its primary Middle Borean subclades, R6 (~51 ka), R30b (~51 ka), and R8 (~42 ka), occur with high frequency among Sri Lankan Vedda as well as Madhya Pradesh Bharia, that is, in two rather distant geographic regions.

In SW Asia, R branches R2'JT (~55 ka), which I would correlate to Pre-Egyptian-Semitic, R0 (~39 ka) correlating to Pre-Basque, and U (~54 ka), and their subsequent subclades. Based on haplogroup frequencies and distinctiveness, I suggest Middle Borean-U subclade U2 correlates to Pre-Dravidian, U8 to Pre-Hattic, U4'9 to Pre-Kartvelian and U3 to Pre-Northwest-Caucasic.

As linguists currently believe Kartvelian emerged in the Holocene, I here give my rationale for the label 'Pre-Kartvelian'. This will also show how I have found it necessary to use the non-linguistic term 'Pre'. The U4'9 homeland is variously postulated as Central Asia/Caucasus/Europe. In my review I found no mtDNA studies identifying U4'9; it appears it has not survived in extant population. There are extensive studies on U4 (~28 ka). Given its frequency cline, the U4 homeland appears to be somewhere along the Eurasian LGM ice-free zone paleolakes, around the southern side of Paleolake Mansi, including the Turgay Spillway to Paleolake Aral and eastward to the headwaters of the Irtysh, Ob, and Yenesei/Angara Rivers, on the Western Siberian Plain. It appears that from there four dispersals occurred:

(a) Northward following the ice retreat along the Yenesei (Ket, presently near confluence of the Tunguska River, and further north Sel'kup and Nenet along the Arctic Kara Sea; Nganasan on the Taymyr Peninsula);
(b) Northward following the ice retreat along the Ob (southern Sel'kup, Mansi, Khant, and northern Mansi along the Arctic Kara Sea) and southward to headwaters of Ob in the Altai (Tubalar)
(c) From Lake Mansi (across the Urals) northward toward the Barents Sea, White Sea, Kola Peninsula area (high frequency U4 in Mesolithic fossil DNA); into the Volga basin and north end of Paleolake Caspian and down to Caucasus (Dargin); and westward into the Dnieper and Danube basins north of Paleolake Black; southward along the eastern Black into Transcaucasia (present day Georgia);
(d) South along the Mansi-Aral spillway and Aral River into Central Asia (Koreimeian Uzbekistan) and the Indus River Valley (Kalash, Pakistanis).

Peoples with high frequency of U4 appear to have adopted languages from various families, including Indo-European (Kalash), Dene-Yeniseian (Ket), Turkic (Tubalar), Uralic Ob-Ugric (Mansi) and Uralic Samoyedic (Nganasan, Nenets). Kartvelian (Swanetians, Georgians) is generally considered an isolate. If one assumes that Nganasan and Nenets peoples adopted a Uralic precursor and innovated Samoyedic, that leaves Kartvelian as the only distinctive language correlation for U4 peoples. So, I tentatively take Kartvelian to be a late innovation on an earlier U4'9 language family, and label that earlier U4'9 family 'Pre-Kartvelian'.

Meanwhile, R in SE Asia/Sahul branches P (~55 ka), which seems strongly associated with Greater-Pama-Nyungan, and subsequently East Asian B4'5 (~50 ka) and Southeast Asian R9 (~47 ka), R21 (~47 ka) and F (~43 ka), the latter three all correlating to Pre-Austric. The dates for P match earliest archaeological dates for Australia ~52 ka.

- Malakunanja II, Kakadu, Northern Territory, Australia, bipolar 'horsehoof' cores, flake lithics, pigments (OSL) 52±8 ka, confirmed; but lowest level artifacts and pigments (OSL) 61±10 ka disputed as disturbed; and Nauwalabila I, Kakadu, lithics, striated ochre pigment, bipolar 'horsehoof' cores (OSL) ~53±5 ka, though disputed (Roberts, Jones and Smith, 1990; Bird, Turney, et al 2002; Flood, 1990; O'Connell and Allen, 2004)

Middle Borean-M [M correlating to speakers of Eurasiatic macrofamily]

- M in East Asia (60.6±13 ka, *SE09*)
  - M → M12'G (57±14 ka East Asia *SE09*; or 47.3±3.6 ka, *BO12)* [Chukotko-Kamchatkan]
  - M → M7 (54.8±13 ka East Asia, *SE09*; or 44.9±3.5 ka, *BO12)* Ryukyu (36%); Ainu (20%); Japan (13%); Korea (11%) [Japonic-Korean-Ainu]; Han Chinese (4-8%); Buryat (5%); Tibet (3%); Kalmyk (2%)
  - M → M80'D → D (48.3±13 ka, East Asia, *SE09*; or 38.4±4.7 ka, *BO12)* → D4 (40.4/34.1 ka, Perego, Achilli et al., 2009) Han Chinese (62% to 32% D depending on location); Tibet (40%); Mongolian Buryat (35%) and Kalmyk (30%); Toto, Sikkim [Tibeto-Burman] (63%)
  - M → M13'46'61 (46.7±3.7 ka, *BO12*) Shannan, southern Tibet (7%M13a,b), Shertukpen, Arunachal Pradesh (23%M61), Lachungpa, Sikkim (12%M61) [all three groups Tibeto-Burman];
  - M → M8 (42.7±12 ka East Asia, *SE09*; or 36.4±7.3 ka, *BO12*)
- M in South Asia (see date for M42'74)
  - M → M42'74 (~55 ka, India/Australia divergence (Kumar, Ravuri, et al., 2009) or 49.6±8.1 ka, *BO12*)
    - M42'74 → M42 (47.7±8.1 ka, *BO12*) → M42b (40.3±7.0 ka, *BO12* or 44.5±12.3 ka, Kumar, Ravuri, et al., 2009) → M42b1a (7.0±6.2 ka, *BO12*) in Madia-Gond, Maharashtra [Dravidian/I.E.]; Munda [Austro-Asiatic]; M74 (35.1±7.2 ka, *BO12*) in Pauri Bhuiya, Orissa [Dravidian/I.E.]
  - M → M33 (44.9±12 ka, *SE09*; 42.3±8.2 ka, *BO12*) multiple India tribes, e.g. Lepcha [Tibeto-Burman] (23%); Mal Paharia, Jharkhand [Austro-Asiatic] (11%); Dungri Bhil, Rajasthan/Gujarat [I.E.] (10%)
  - M → M5 (40±12 ka, *SE09*; 37.1±14.8 ka, *BO12*) multiple India tribes, e.g. Dungri Bhil (25%); Andh, Maharashtra, Andhra Pradesh, Chhattisgarh [I.E.] (18%); Kamar, Chhattisgarh [Dravidian] (15%); Nihal [Nihali isolate/I.E.] (8%)
  - M → M2 (38±11 ka, SE09; 36.3±6.1 ka, BO12) Betta Kuruba, Karnataka [Dravidian] (64%); Toda, Nilgiri Plateau, Tamil Nadu [Dravidian] (44%); Korku, Central India [Austro-Asiatic] (37%); Madia-Gond, Maharashtra [Dravidian/I.E.] (32%); Katkari, Maharashtra & Madhya-Pradesh [I.E.] (31%); Hill Kolam, Central India [Dravidian/I.E.] (29%); Ka Thakar and Ma Thakar Maharashtra [I.E.] each (29%)
  - M → M4"67 (34.5±4.2 ka, *BO12*)

- M in SE Asia/Sahul (49.4±10 ka, *SE09*; or average ~64 ka, Jinam, Hong et al., 2012)
  - M → M9 (53.4±15 ka, East or SE Asia/Sunda, *SE09*, or 38.4±8.4 ka, *BO12*) → M9ab'E (46.9±16 ka) → M9a'b (33.4±13 ka *SE09*; or 27.2±8.1 ka, *BO12*) [Sinitic, Tibeto-Burman, Tai-Kadai, Hmong-Mien, Austronesian] and E (27.4±11 ka, SE Asia) [Papuan, Austronesian]
  - M → M29'Q (46.6±3.8 ka, *BO12*) → Q (37.5±5.6 ka, *BO12*) Melanesia, Australia → Q2 (28.7±7.3 ka, *BO12*) Baining, East New Britain (35%), Ata, West New Britain (29%) [both Papuan]
  - M → M21 (45.6±4.6 ka, *BO12*) South Asia, Bangladesh, SE Asia; Semang Negritos [Aslian Austro-Asiatic] (Mendriq 87%M21a,b; Batek 48%M21a);
  - M → M27 (45.3±3.5 ka, *BO12*) Melanesia, Bougainville, New Britain, New Ireland
  - M → M22 (40.6±5.6 ka, *BO12*) Temuan Aboriginal Malay [Malayan, Austronesian] (range 18-6% M22,a + 9-28%M21a,b)
  - M → M42'74 (see above) → M42a (40.6±14.8, Kumar, Ravuri, et al., 2009; or 33.2 ka, van Holst Pellekaan, et al., 2006) only in Australia and Tasmania; Paakintji, Lower Darling River, SE Australia [Pama-Nyungan] (40%)

As noted above, the Early M dispersal appears to begin from East Asia, from there to South Asia, and later into SE Asia/Sahul, and taken about 12,000 years. According to the Middle Borean mtDNA branches and their TMRCAs, it appears that M12'G-mtDNA is strongly associated Chukotko-Kamchatkan and M7 with Japonic-Korean-Ainu. Since D-mtDNA is a branch of M, one might predict populations with high frequencies of D to be speakers of a language in the Eurasiatic family. Unexpectedly high frequency of D is found in Han Chinese speakers, and thus I infer that the ancestors of the Han originally spoke a Eurasiatic language but at some later time abandoned it and adopted or developed Chinese from a Borean-N language in the Dené-Caucasian family. The M13'46'61 branch seems to correlate to Pre-Tibeto-Burman; M29'Q to Pre-Papuan; and M8 (CZ→C) to Pre-Altaic-Mongolian-Tungusic and the Yukaghir isolate. M in South Asia, which is associated with current Austro-Asiatic speakers, appears to be a back-migration from SE Asia.

Thus, by around 50,000 years ago (MIS 3c/b), based on mtDNA phylotree TMRCAs, it appears that roughly 28 new haplogroups had emerged, and based on current languages associated with them, I suggest that at least 17 of these haplogroups correlate to the emergence of 17 language families from Europe to East Asia and the Sahul. I suggest the term 'Middle Borean' to designate the ancestors of these language families. Keeping in mind the caveat that the date for the emergence of a particular protolanguage may likely will be later than the mtDNA TMRCA, the genetic and language family correlations I hypothesize at around 50 ka are summarized in Table 7 and Figure 6. Figure 6 circles are placed on the map to illustrate approximate homelands; original mtDNA homelands, migration routes and destinations (current homelands) may have covered wider or more irregular regions.

Table 7: ~50 ka (MIS 3c/b) 'Middle Borean'			
Early Borean-N	~71 ka (S Asia); ~62 ka (W Eurasia); ~58 ka (E Asia); ~56 ka (SE Asia/Sahul) *Dené-Caucasic (Burushaski)*		
Middle Borean-N			root-N, Pestera cu Oase, ~40ka
N1'5	~57 ka		
N1	~52 ka	Pre-Burushaski	macro-N in Gravettian Paglicci12, ~29 ka
N2	~44 ka	Pre-Northeast Caucasic	
N9, Y	~49 ka	Pre-Nivkh (adopts Borean-M Pre-Eurasiatic language?)	
N11	~56 ka	in SE Asia (adopts Austric?)	
S	~53 ka	(only Australia, Tasmania; adopt non-Pama-Nyungan & P-N	
O	~48 ka	(adopts Pama-Nyungan)	fossil O1a, SW Australia 100BP
Early Borean-M	~61 ka (E Asia); ~55 ka (S Asia); ~49 ka (SE Asia/Sahul) *Eurasiatic*		
Middle Borean-M			
M12'G	~57 ka	Pre-Chukotko-Kamchatkan	
M7	~55 ka	Pre-Japonic-Korean-Ainu	
M9	~53 ka	(adopt Tibeto-Burman, Sinitic, Austric, Papuan)	
M42'74	~52 ka	(adopts various Dravidian, Austroasiatic, I.E.)	
D	~48 ka	(adopts Borean-N Dené-Caucasic-Burushaski as Sino-Tibetan?)	
M13'46'61	~47 ka	Pre-Tibeto-Burman	
M29'Q (Q2)	~47 ka	Pre-Papuan	
M9ab'E	~47 ka	(adopt Sinitic, T-B, Austric, Papuan)	
M21	~46 ka	(adopts? Aslian Austro-Asiatic)	
M33	~45 ka	(adopts various Tibeto-Burman, Austroasiatic, I.E.)	
M27	~45 ka	(adopts Papuan?)	
M8	~43 ka	Pre-Altaic-Mongolian-Tungusic	
M22	~41 ka	(adopts Malayan, Austronesian)	
M42a	~41 ka	(adopts Pama-Nyungan)	
M5, M2	~40 ka	(adopt various Dravidian, Austroasiatic, I.E.) M5 in Nihal	

Early Borean-R	~67 ka (S Asia); ~59 ka (W Eurasia); ~58 ka (SE Asia/Sahul); ~54 ka (E Asia) *Afroasiatic, Dravidian, Hattic, Austric, P-N*		
Middle Borean-R South Asian			
R6	~51 ka	Bharia (adopts Dravidian)	
R30b	~51 ka	Vedda substrate	
R8	~42 ka	in Vedda	
Middle Borean-R Western Eurasian			root-R in fossil Ust-Ishim, ~45 ka; R in fossil Fumane, ~41 ka
R2'JT	~55 ka		
→JT	~50 ka	Pre-Semitic-Egyptian	
R→U	~54 ka		basal U in fossil Mal'ta MA1, ~24 ka
U2	~54 ka	Pre-Dravidian	fossil U2 in Aurignacian Kostenki14, ~38 ka
U8	~50 ka	Pre-Hattic-Kaskian?	fossil U8 in Gravettian Dolní Vestonice13, ~31 ka
U4'9	~43 ka	Pre-Kartvelian	
U3	~41 ka	Pre-Northwest-Caucasic	
R0	~39 ka	Pre-Basque (also in Marsh Arabs, Iraq, presumed descendents of Sumerians)	fossil R0/HV in Gravettian Paglicci25, ~27 ka
U6	~36 ka	(back migration into Northern Africa, adopts Berber)	Early to Late UP Dabban 43 to 17 ka, Haua Fteah
Middle Borean-R SE and E Asian			
P	~55 ka	Greater-Pama-Nyungan	
B4'5	~50 ka	Pre-Austric	fossil B* in Tianyuan, ~40 ka
B4	~44 ka	(Austronesian, Austroasiatic, adopts Trans-New-Guinea)	
B5	~42 ka	(Kam-Tai, Tibeto-Burman)	
R9	~47 ka	Pre-Austric (Malay Aslian Austroasiatic, Kam-Tai, Austronesian, Hmong-Mien)	
R21	~47 ka	Pre-Austric (Aslian Austroasiatic)	
R9→F	~43 ka	Pre-Austric (in Mon-Khmer, Austroasiatic, Tai, Aslian)	

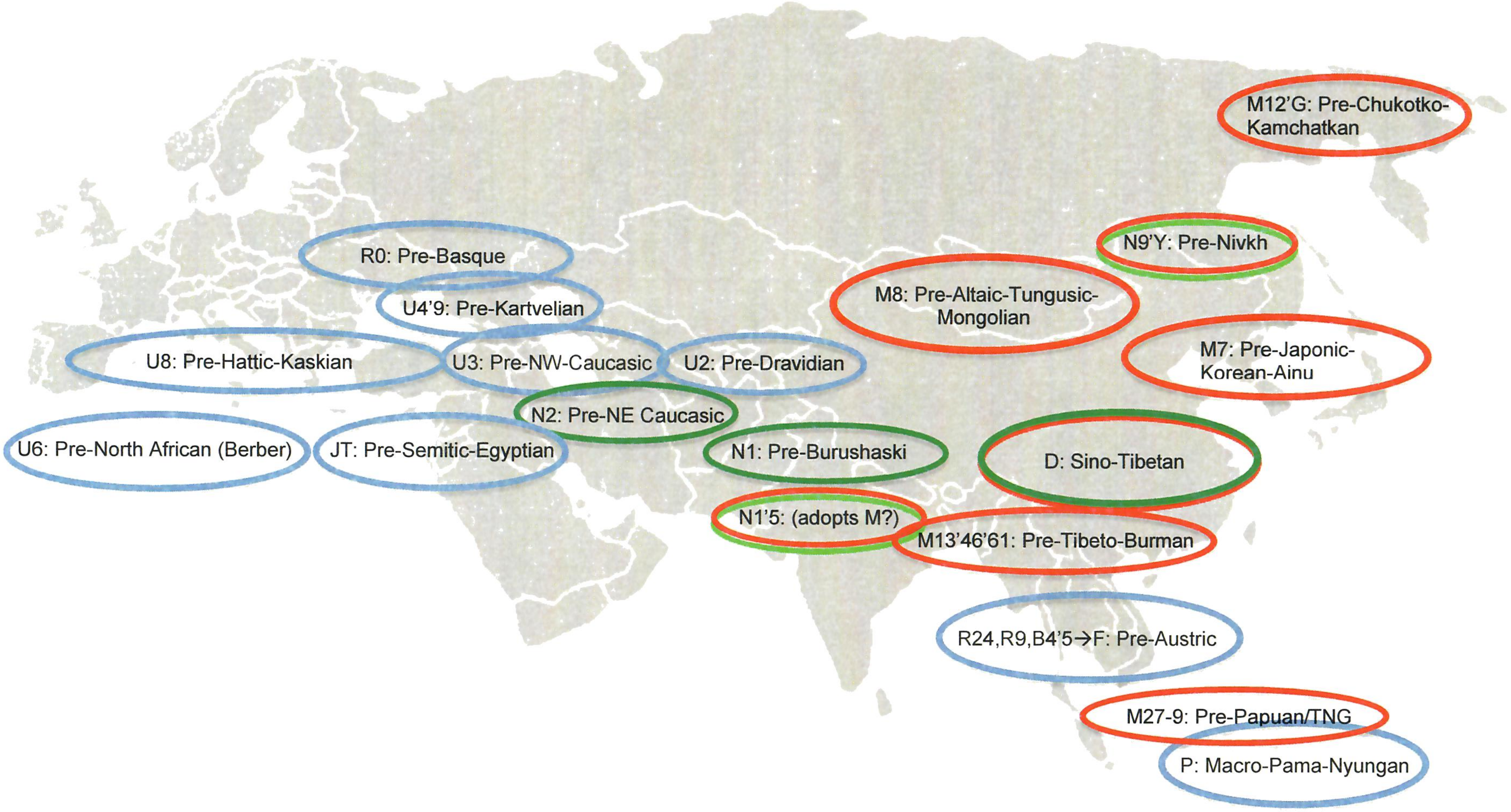

M12'G: Pre-Chukotko-Kamchatkan
N9'Y: Pre-Nivkh
R0: Pre-Basque
M8: Pre-Altaic-Tungusic-Mongolian
U4'9: Pre-Kartvelian
M7: Pre-Japonic-Korean-Ainu
U8: Pre-Hattic-Kaskian
U3: Pre-NW-Caucasic
U2: Pre-Dravidian
N2: Pre-NE Caucasic
U6: Pre-North African (Berber)
JT: Pre-Semitic-Egyptian
N1: Pre-Burushaski
D: Sino-Tibetan
N1'5: (adopts M?)
M13'46'61: Pre-Tibeto-Burman
R24,R9,B4'5→F: Pre-Austric
M27-9: Pre-Papuan/TNG
P: Macro-Pama-Nyungan

Figure 6. Out-of-Africa mtDNA Haplogroups Ancestral ('Pre') to Language Macrofamilies at circa 50,000 Years Ago
Blue: Borean-R (Pre-Afroasiatic, etc.); Green: Borean-N (Pre-Caucasic-Burushaski-Dené); Red: Borean-M (Pre-Eurasiatic)

M, N and R-mtDNA: Late Borean. By around 25,000 years ago another dozen language families emerged—which I term 'Late Borean'—and this, for the most part, appears to have completed the development of the major language families of the world.

Late Borean-R [R correlating to speakers of Afroasiatic, Kartvelian, Dravidian, Elamitic, Austric and Pama-Nyungan families]

- R in W Eurasia
  - JT → J (32.6±11 ka, *SE09*; 34.3±4.9 ka, *BO12*; 34.5/43.3 ka, Pala, Olivieri, et al., 2012) Saudi Arabia (21%); SW Asia general (12%), Iran (13%); Europe (11%); Egypt (9%); Basque; Caucasus (8%); South Asia
    - J → J1 (24.1±8.5 ka, *SE09*; 26.9±5.3 ka, *BO12*; 25.0/33.3 ka, Pala, Olivieri, et al., 2012) Eastern Europe, Ukraine, Balkans, arrived post-LGM; in Neolithic and Mesolithic fossil DNA
    - J → J2 (28.3±4.6 ka, *BO12*; 32.9/36.8 ka, Pala, Olivieri, et al., 2012) SW Asia; J2a1 and J2b1 in Europe ~15-16 ka (Pala, Olivieri, et al., 2012). (Fossil J or less likely JT in Solutrean, Nerja, Málaga, ~20-24 cal ka (Fernández, thesis, 2005) [JBH: Oven & Kayser Build 14 has some Nerja SNPs in JT, J1 and J2; and Pala et al., (2012) has 3 SNPs for J2b1 similar to Nerja fossils.]
  - R0 → HV (27.1±7.5 ka, *SE09*; 21.9±2.8 ka, *BO12*; Near East) Syria (24%R0s/HV+ 16Hs), Basque (16%HV+52%H+7%V or 3%to14%HV+43%to67%Hs); Iraqis (13%HVs+ 4%R0a+16%Hs), Persians (2%R0+11%HVs+17%Hs); Dargin (9%HVs+24%Hs) and Avar (8%HVs+23%Hs) [NE Caucasian]; Marsh Arabs (8%R0/R0a+4%HVs+12%Hs) [Sumerian?]; fossil Minoan (3%R0+8%HV+32%Hs)
    - HV → HV0 ('pre-V') (19±7 ka, *SE09*; 13.5±3.2 ka, *BO12*; before LGM, perhaps Eastern Europe, spreading E to W along Gravettian axis); HV0a, HV4a1 Basque
    - HV → H (18.6±4 ka, *SE09*; 12.8±0.8 ka, *BO12*) Basque (16%HV+52%H+7%V or 3%to14%HV+43%to67%Hs); Scandinavia (49%); Sicily (49%), Germany (49%), France (47%), North Italy (47%); European Russians (42%); Dargin (24%), Chechen (24%), Avar (23%). (Fossil H in Magdalenian La Pasiega; in Mesolithic Villabruna, ~14 cal ka.)
  - U → U1 (37±11 ka, *SE09*; 32.0±5.4 ka, *BO12*) Kubachi [NE Caucasian] (48%), Azeri [Turkic] (8%), Lur Zagros Iran [IE] (6%), Persians [IE] 3%. [Pre-Elamitic?]
  - U → U5 (36±11 ka, *SE09*; 30.2±5.3 ka, *BO12*) Finno-Lapic Saami (48%U5b1b1), Seto (23%), Finns (19%), and Karelians (17%); Finno-Permic Mordvin (16%) European Russia [I.E.] (14%); Basque (11%); Hungarian (8%). (Fossil DNA, U5 mutations in Gravettian Dolní Vestonice 14 and 15, ~31 cal ka; U5 n Magdalenian Cantabria; U5b1 in Late UP Federmesser, ~14 cal ka.) [Pre-Finno-Ugric/Uralic]

  - U8→U8b (35±12 ka, *SE09*; 38.4±4.9 ka, *BO12*) Kurdish, W Iran [IE] (10%); Qatar (10%); U.A.E (8%); Gilaki, N Iran, SW Caspian [IE] (2.7%); Konya, Anatolia [Turkish] (2.0%); Jordan, Italy; U8→U8a (37±14 ka, *SE09*; 18.5±5.2 ka, *BO12*) SW Europe (2.8%); Basque (1.1%); north-central Europe (0.2%), Anatolia (0.2%); Basque since 28±9 ka, U8a in Basque expansion 23±14 ka, U8a1 expansion 14±5 ka;
    - U8b→K (31.9±11 ka, *SE09*; 26.7±4.3 ka, *BO12*) Druze (16%); Swanetia [Kartvelian] (12.5%); Kurdish Turkmenistan [IE] (12.5%); Georgian (10%); Kurdish, Iran [IE] (10%); U.A.E. (10%); Caucasus (8.6%); Palestine, and Iran (8%); Syria, Iraq and Lur Zagros Iran (6%); Turkey (5%); Jordan (4%); Hazara [IE]; Baluch [IE]; Europe [IE], Basque (K1a1), Ashkenazi (K1a1b1a). (Fossil DNA, K-mtDNA in LBK Germany (14%); is a Neolithic marker.)
  - U → U7 (22±10 ka, *SE09*, homeland SW Asia between Gujarat and Iran; 18.1±3.7 ka, *BO12*) Iran Kurds [I.E.] (20%); Gilaki, SW Caspian [I.E.] (11%); Brahui [Dravidian] (10.5%); Gujarati [I.E.] (9%); Sindhi [I.E.] (9%); Hunza [Burushaski] (7%);
  - U4'9 → U4 (21±10 ka, *SE09*; 17.5±3.1 ka, *BO12*; 28.3±9.3 ka, Malyarchuk, 2004) Kalash [I.E.] (34%), Ket [Dené-Yeniseian] (29%), Nganasan [Samoyedic] (21%), Tundra Nenets [Samoyedic] (13%'U'), Tubalar [Altai] (15-18%), Western Siberia (Mansi, Nentsi, Nganasan, Ket average) (17%), Mansi [Ob-Ugric] (16%), Pakistan (14%), Volga-Ural peoples (9.7%), Hazara [I.E.] (9%), Swanetia (8.3%) and Georgians [Kartvelian] (8%). (Fossil U4, Mesolithic Lake Onega cemetery, ~8 cal ka.)
- R in SE Asia, East Asia, Beringia and Americas
  - B4'5 → B4 → B4b (28±9 ka, *SE09*), Tubalar (6%), Mongol (15%B4), W. Evenk (4%B4), Altai-Kizhi, Tuva, Tofalar
  - B4'5 → B4 → B4a (26±8 ka, *SE09*) China, Thailand, Indonesia, Taiwan, Philippines [Austronesian], B4a1a (10±5 ka, *SE09*) 'Polynesian motif'; Karkar Islanders [Austronesian and Trans-New-Guinea] (11%B4a+23%B4a1a1);
  - B4'5 → B5 → B5a (27.6±5.9 ka, *BO12*) Musuo, Yunnan [Na<Tibeto-Burman] (13%B4+17.4%B5a]
  - B4'5 → B4 → B2 (21.2±2.4 ka, Achilli, Perego, et al., 2008) in North American tribes such as Kiliwa (Baja) [Yuman] (100%); Jemez Pueblo [Tanoan} (89%); Zuni [isolate] (77%); Penutian Wintuan, Utian, Miwok, Costanoan, Yokut (56%) and Salish-Sahaptian-Yakama-Wishram (50-67%); South American tribes such as Matsiguenga, Peru [Arawak] (92%); Aché, Paraguay [Guarani<Tupian] (90%); Xavánte, Mata Grosso [Ge] (84%); Uros [Uruquilla isolate] (73%); Aymara [Greenberg: Andean] (72%); Quechua [Andean] (61%) [Question: does this grouping correspond to Y. Berezkin (2010a, 2010b) 'Indo-Pacific' in the Americas?]

Late Borean-N [N correlating to Burushaski-Caucasic-Dené macrofamily]

- N in East Asia, Beringia and Americas
  - N → A (29.2±10 ka, *SE09*; Caspian-Baikal homeland; or 24.2±4.9 ka, *BO12*)
  - A → A2 (15.6±1.8 ka, *BO12*; or 18.1±1.8 ka, Achilli, Perego, et al., 2008); Na Dené (95%); Dogrib, Tlingit (100%); Haida (85-96%); Maya , Quintana Roo [Mayan] (80%); Bella Coola [Salishan] (65%); Mixtec-Alta [Otomanguean} (73%); Mixe [Mixe-Zoquean] (63%); Apache (63%); Navaho [Dené] (52-58%A2+38-41%B); Chumash (53-60%); Nahua (57%A2+28%B2) and Huichol (56%A2+25%B2) [both Uto-Aztecan]; Kuna (100%), Ijka (90%), Arsario (68%), Kogi (65%) [all four Chibchan-Paezan]; Guarani, Brazil [Tupian] (84%); Barasano [Equatorial-Tucanoan] (80%); Kaingang [Ge] (62%) [I would call this 'Pre-Dené' or 'A-Amerind']
  - N9'Y → Y (22±11 ka, *SE09*; or 24.6±7.1 ka, *BO12*) Nivkh [isolate] (66%); Ulchi SE Tungusic] (43%+4%N9b); Ainu [isolate or Japonic-Korean-Ainu] (20%+8%N9b)
- N in W Eurasia
  - N1→N1ae'I → N1e'I → I (26.3±10 ka, *SE09*) Northern Europe (2-4%), Ukraine (11%), Pakistan (9%) [I.E.], El Molo [Cushitic] (22%); N1a (19.3-22.3 ka, Fernandes, Alshamali et al., 2012), with branches in Central Asia, SW Asia, Europe, in LBK Neolithic fossil
  - N2 → N2a (24±8 ka, *BO12*) Caucasus, Eastern Europe;
  - N2→W (21±8 ka, *SE09*) Indus Valley, SW Asia, Europe, NW Africa, India; Neolithic marker; Sindhi [I.E.] (17%); Finns [I.E.] (10%); Kurds [I.E.] (10%); Mazandar, Iran [I.E.] (9.5%); Gujarati [I.E.] (9%); Swanetia [Kartvelian] (8%)
  - N → X (31±13 ka, *SE09*) Avar (15%X [basal X]+6%N) → X1'2'3 (~28.8 ka, Fernandes, Alshamali et al., 2012) → X1'3 (~21 ka, Fernandes, Alshamali et al., 2012) Druze (16%X1'3+11%X2); Georgian (8%X) and → X2 (20.9±9 ka, *SE09*; 19.2±2.6 ka, *BO12*) populations across SW Asia, Europe, North Africa, Central Asia, Siberia
    - X2 → X2a'j (17.1±3.1 ka, *BO12;* ~19.4 ka, Fernandes, Alshamali et al., 2012)
    - → X2a (12.8±7 ka, *SE09*; 12.7±3.5 ka, *BO12*; ~14.1 ka, Fernandes, Alshamali et al., 2012) Micmaw (50%X+33%A), Anishinabe/Ojibwa, Minnesota (50%), Cheyenne (18%+50%A) [all three Algonquian]; Nuu-chah-nulth (7%) and Yakama (5%) [both Penutian]; Navaho [Na-Dené] (3%) ['X-Amerind']. (Fossil X2a in 'Kennewick Man', ~8.5 cal ka.)
- N in South Asia
  - N1'5 → N5 (37±8 ka, *BO12*) India, Madhya Pradesh Sahariya [I.E.]
- N in SE Asia/Sahul
  - N → N22 (25.2±8.8 ka, *BO12*) Temuan Aboriginal Malay [Malayan, Austronesian] (12%N22+15%N21 or17%N22+22%N21) and →N21 (22.4±9.0 ka, *BO12*) Semelai Aboriginal Malay (31%N21); Temiar Senoi (31%N21). M21 in Temuan appears derived from ancestral type found in Cham of Vietnam (Jinam, Hong, et al., 2012)

  - N → N12,13,14 (~17 ka, Hudjashov, Kivisild et al., 2007) in Kalumburu, others Australia [Pama-Nyungan and non-Pama-Nyungan speakers]
  - N → S → S2 (38±7 ka, *BO12*; or 15.2±5.1, Hudjashov, Kivisild et al., 2007) and → S1 (22.0±7.7 ka, Hudjashov, Kivisild et al., 2007) Paakintji, Lower Darling River (7%S1+1%S2) and Yuendumu-Warlpiri, Central Desert (10%S1+8%S2+1%S5) [both Pama-Nyungan] and Kalumburu, Kimberly [non-Pama-Nyungan] (28%S2+6%S5)

Late Borean-M [M correlating to speakers of Eurasiatic macrofamily]

- M in East Asia, Beringia and Americas
  - D4 → D1 (16.8±2.9 ka, *BO12*; 16.9±1.6 ka, Bodner, Perego, et al., 2012) → D1g (11.6±4.4 ka, *BO12*) Chile and Argentina, esp. Mapuche, thus earliest stage of rapid coastal route, Beringia to Southern Cone dispersing in less than 2,000 years (Bodner, Perego et al., 2012) Yahgan [Yahgan/Yamana isolate] (10%D1+33%D1g2+10%D4h3a); Huilliche, central Chile (4%D1+37%D1g+4%D4h3a) and Mapuche, central Argentina (4%D1+24%D1g+1%D1j) [both Araucanian isolate<Andean]; Mayo, Sinaloa [Uto-Aztecan] (33%); Kawésqar [Alacalufan isolate], nomadic seafarers (8%);
    D (subclade not specified): Shoshone/N. Paiute [Uto-Aztecan] (48%);Yok-Utian [Penutian] (47%); Wapishana, Brazil (67%), Zoró (60%) and Tucano (57%) [all three Equatorial-Tucanoan]
  - D4 → D4h (21.5±2.4 ka, *BO12)* → D4h3 (18.3±2.9 ka, *BO12)* → D4h3a (13.0±2.6 ka, *BO12)* mostly South America, Chile, less so in Mexico, California, confirms a coastal route (Perego, Achilli et al., 2009) Kawésqar [Alacalufan isolate], nomadic seafarers (46%); Tehuelche, central southern Argentina [Chonan] (28%D4h3a+16%D1g); Cayapa, Ecuador (22%); Chumash (16%); Aonikenk, Fuego [Chonan] (73%D). (Fossil D4h3a in Anzick-1, ~12.5 cal ka.)
  - M12'G → G (35.7±10 ka East Asia *SE09*; or 31.3±5.7 ka, *BO12)* Itelman (68%) and Koryak (42%) [both Chukotko-Kamchatkan]
  - M8 → CZ → Z (24±9 ka Caspian-Baikal, *SE09*; or 21.7±8.4 ka, *BO12*) and CZ → C (28±9 ka Caspian-Baikal, *SE09*; or 23.9±4.8 ka, *BO12*) Yukaghir [isolate language] (66-72%); E. Evenk (62%), W. Evenk (50%) and Even (43%) [all three N. Tungusic]; Tofalar (61%), Tuvan (51%) and Yakut (47%) [all three Turkic]; Nganasan (51%), Tundra Nenets (32%) and Sel'kup (22%) [all three Samoyedic]; Altai-Kizhi [Altaic] (34%); Mongol (19%)
  - M8 → M8a (26±10 ka, China, Japan, *SE09*), Dirang Monpa, Arunachal Pradesh [Tibeto-Burman] (24%)
  - C → C1 (17.1±5 ka, *SE09*; 18.3±4.2 kya, *BO12*) Ayoreo [Zamucoan] (84-100%); Ancient Maya-Copan (89%); Baja Seri [isolate] (88%); Pima (82%) and Tarahumare

(54%) [both Uto-Aztecan]; Taino [Arawakan] (75%); Makiritare [Ge-Pano-Carib] (70%); Movina [Equatorial-Tucanoan] (64%); Yanomama [Chibchan-Paezan] (54-72%); Mapuche, central Argentina [Araucanian isolate<Andean] (42%)

- M in South Asia
  - M → M32'56, M31, M39'70, M36, M49, M3, ~35 ka; and M60, M6, M19'53, M41, M44, <30 ka [associated with/adopt various language families]
- M in SE Asia/Sahul
  - M → M75, M21a,b, M76, M45, M26, M51, M20, M1, M9a,b, E; all <30 ka [associated with/adopt various language families]
  - M29'Q → Q (37.5±5.6 ka, *BO12*) Melanesia, Australia → Q1 (18.2±7.3 ka, *BO12*) Muyu, Irian Jaya Highlands [Ok] (88%), Lowland Riverine (Mandobo), West Papua (85%); Asmat, Irian Jaya SW Coast [Asmat] (84%), Dani, Irian Jaya Highlands [Dani] (76%), Una (pygmoid), Irian Jaya Highlands [Mek] (62%); Bandi [Chimbu-Wangi} (29%) [all Trans-New-Guinea]; Aita, North Bougainville [Papuan] (82%)

Thus, by around 25,000 years ago, based on mtDNA phylotree TMRCAs, it appears that roughly two dozen new haplogroups had emerged, and based on current languages associated with them, I suggest that at least a dozen of these haplogroups correlate to the emergence of distinct language families in Europe, East Asia and the Sahul. I suggest the term 'Late Borean' to designate the ancestors of these language families. Keeping in mind the caveat that the date for the emergence of a particular protolanguage may likely will be later than the mtDNA TMRCA, the genetic and language family correlations I hypothesize at around 25 ka are summarized in Table 8.

Table 8: ~25 ka (MIS 2) ‘Late Borean’			
Late Borean-N			
A	~29 ka	(Caspian-Baikal)	
A2	~16 ka	A-Amerind (Pre-Dené)	
W	~21 ka	(adopts I.E.)	
Y	~22 ka	Nivkh (adopts Pre-Eurasiatic language?)	
X, X1’2’3	~31 ka	(SW Asia, Central Asia, Siberia)	
X2a’j X2a	~17 ka ~13 ka	X-Amerind (Algonquian)	fossil X2a in Kennewick, ~8.5 ka
N22	~25 ka	(in Aboriginal Malay, Senoi)	
N12, N13, N14		(in Pama-Nyungan and non-Pama-Nyungan)	
S1 S2	~22 ka 38/15 ka	(in Pama-Nyungan and non-Pama-Nyungan)	
Late Borean-M			
G	~35 ka	Chukoto-Kamchatkan	
C	~28 ka	Pre-Altaic-Tungusic-Mongolian and Yukaghir (isolate)	
E	~23 ka	(adopts Papuan, Austronesian)	
Q1	~18 ka	Trans-New-Guinea	
D1	~17 ka	D-Amerind (wave 1?)	
D4h3a	~13 ka	D-Amerind (wave 2?)	fossil D4h3a in Anzick-1, 12.5 ka
C1	~17 ka	C-Amerind	
M9a	~15 ka	(adopts Tibeto-Burman)	
M9b	~19/3 ka	(adopts various Austric)	

Late Borean-R Western Eurasian			
U1	~37 ka	Pre-Elamitic (?)	
U5	~36 ka	Pre-Finno-Ugric	fossil U5 mutations in Gravettian Dolní Vestonice14 &15, ~31 ka; U5b1 in Late UP Federmesser, ~14 ka; U5 in Cantabria, Magdalenian ~14 ka
U8→U8b	~35 ka	(in SW Asia, Iran, Qatar, UAE, Anatolia, Italy)	
U8b→K	~30 ka	(in SW Asia, Druze, Kurds, Caucasus, Georgia, Iran, Turkey, Europe)	LBK Germany (14%) Neolithic marker
U8→U8a	~37 or 18 ka	(in SW and N Cen Europe, Basque, Anatolia)	
J	~33 ka	(in SW Asia, Europe, S Asia)	
→J2	~28 ka	(in SW Asia, Europe)	fossil J2 in Solutrean Nerja, ~20-24 ka
HV	~27 ka	(in SW Asia, Caucasus, Basque)	fossil RO or HV in Gravettian Paglicci 25, ~23 ka
→HV subclades	~19 ka	(in Basque, Saami, E. Eurasia)	
→H	~19b ka	(in Basque, E. Eurasia)	fossil H in Magdalenian La Pasiega; in Mesolithic Villabruna, ~14 ka
U7	~22 ka	Pre-Harappan	
U4	~21 ka	(in Nganasan, Nenets, Ket, Kalash, Mansi, Dargin, Swanetia, Georgians)	fossil U4, Mesolithic Onega, ~8 ka
Late Borean-R SE Asian, E Asian Beringia, Americas			
B4b	~28 ka	(in Mongolia, Siberia, various languages)	
B5a	~28 ka	(in Tai-Kadai, Tibeto-Burman)	
B4a	~26 ka	Austronesian	
B4a1a	~10 ka	(Polynesian motif)	
B2	~21 ka	B-Amerind	'Indo-Pacific' in Americas (Berezkin)

Discussion.

This discussion section focuses on out-of-Africa geographic bottlenecks and derives correlations of mtDNA and language family dispersals for each bottleneck. To provide a detailed discussion of each geographic region of the world is beyond the scope of this paper. For this level of detail the reader is encouraged to peruse the Master Database (Supplementary File, Table 1, https://originsnet.academia.edu/JamesHarrod), which to a large extent is chronologically ordered and ordered by region. In addition Appendix 1 consists of out-of-Africa mtDNA and correlated language family dispersal maps. Five maps are posited for each of the major stages of mtDNA phylotree haplogroup evolution and TMRCA dates, illustrating apparent dispersal routes and correlations to language families at each stage. In drawing these maps I superimposed my dispersal hypothesis onto maps illustrated in Mishra, Chauhan & Singhvi (2013: Fig. 2), for which I am grateful.

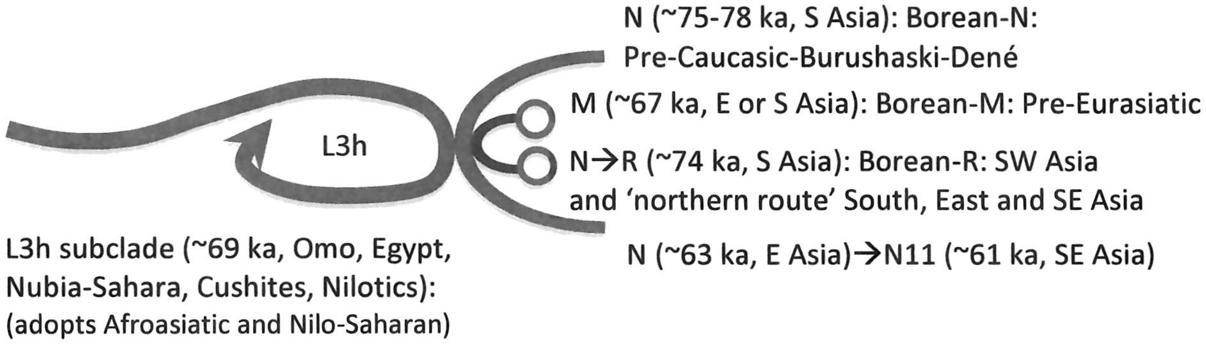
Based on the results of my meta-pattern-analysis it appears that basic mtDNA haplogroup and language family emergences and differentiations occur each of seven (7) major geographic bottlenecks that had to be crossed from Africa through Asia to Sahul and the Americas.

- Africa to SW Asia Crossing (circa 75 ka, differentiating L3 and Early Borean-N, R and M)
- SW Asia to S Asia, the Zagros Crossing (circa 65 ka, differentiating Middle Borean-R
- Ganges Delta Crossing to SE Asia (circa 50 ka, differentiating Middle Borean-R and Middle Borean-M)
- SW Asia to W Eurasia, the Transcaucasus. Crossing (circa 50 ka, differentiating Middle Borean-U)
- 'Silk Route Crossing', Persian and Indus Spurs bifurcation west to Europe, east to Central and East Asia (circa 25 ka, differentiating Late Borean-N)
- Sunda/Sahul Crossing (circa 50 ka, differentiating Middle Borean-N, M and R)
- Siberia to the Americas, the Beringia Crossing (circa 25 ka, differentiating Late Borean-N, R and M)

The mtDNA phylotree branchings at these seven bottlenecks seem to display a common pattern. On the proximal side of a bottleneck, one or more haplogroups seem to swirl backward in the opposite direction from the narrow passage; on the distal side, haplogroups spurt out in divergent directions. At E-W crossings (Africa-SW Asia, Zagros and Ganges), they diverge in northerly, central and southerly directions. At N-S crossings (Transcaucasus and Silk Route), they diverge in easterly, central and westerly directions. They can be visualized schematically as having the shape of a scorpion, and so I call them 'scorpion bottlenecks'.

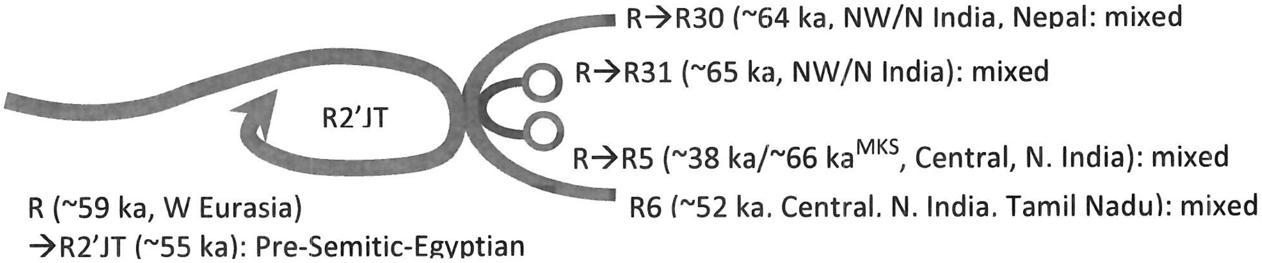
Based on my meta-pattern-analysis results I diagram eight (8) correlations of mtDNA haplogroups and language families at the seven geographic bottleneck crossings. TMRCA dates and original or current homelands are those identified in the results section of this study.

1. Africa to SW Asia Crossing (circa 75 ka, differentiating L3 and Early Borean-N, R and M). (TMRCA dates for this crossing are multiplied 1.1x to account for earlier revised CHCLA date, see narrative.)


N (~75-78 ka, S Asia): Borean-N:
Pre-Caucasic-Burushaski-Dené
M (~67 ka, E or S Asia): Borean-M: Pre-Eurasiatic
L3h
N→R (~74 ka, S Asia): Borean-R: SW Asia
and 'northern route' South, East and SE Asia
L3h subclade (~69 ka, Omo, Egypt,
Nubia-Sahara, Cushites, Nilotics):
(adopts Afroasiatic and Nilo-Saharan)
N (~63 ka, E Asia)→N11 (~61 ka, SE Asia)

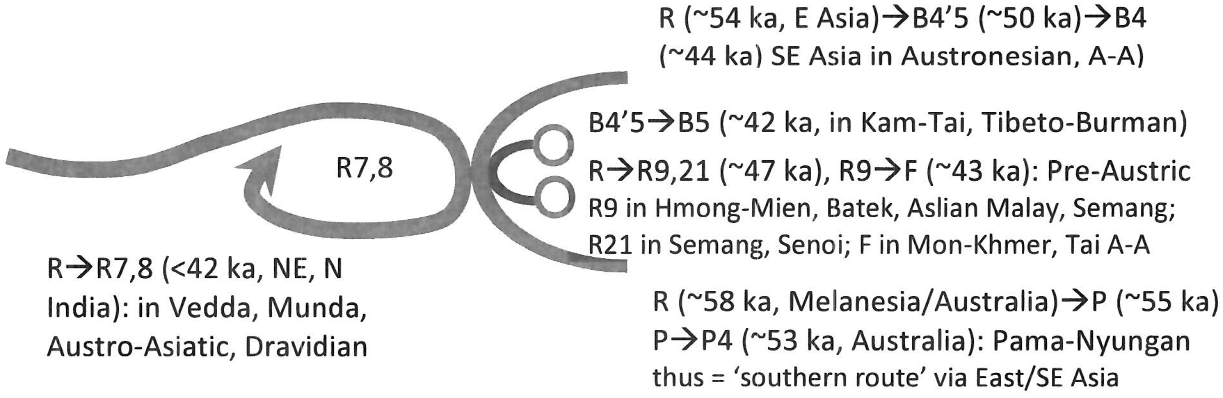
L3 dates around 80 ka. Distal branches N and M occur only out-of-Africa in SW Asia, with N dating about ~75-78 ka; R-mtDNA from N at around 74 ka; and M somewhat later at ~67 ka. N arrives in E Asia around 63 ka and in SE Asia a couple thousand years later. There is insufficient well-dated archaeology to evaluate whether the L3 dispersal route into SW Asia was via the Sinai, the Bab el Mandeb or around the Red Sea, or some combination of these. The scorpion diagram suggests taking as its proximal 'backward movement' the oldest L3 subclade, L3h, which has a TMRCA later than N and which has highest frequencies among Egyptians, Nubian-Saharans. Omotics and Nilotics. As noted earlier, these dates suggest that the *sapiens sapiens* dispersal out-of-Africa endured a roughly 5,000-year pause in SW Asia. A *sapiens sapiens* 'fast track' on some sort of 'southern route' sometime between 60 to 45 ka does not appear supported.

2. SW Asia to S Asia, the Zagros Crossing (circa 65 ka, differentiating Middle Borean-R). ('mixed'= Dravidian, Indo-European, Austro-Asiatic)


R→R30 (~64 ka, NW/N India, Nepal: mixed
R→R31 (~65 ka, NW/N India): mixed
R2'JT
R→R5 (~38 ka/~66 ka^MKS, Central, N. India): mixed
R (~59 ka, W Eurasia)
→R2'JT (~55 ka): Pre-Semitic-Egyptian
R6 (~52 ka, Central, N. India, Tamil Nadu): mixed

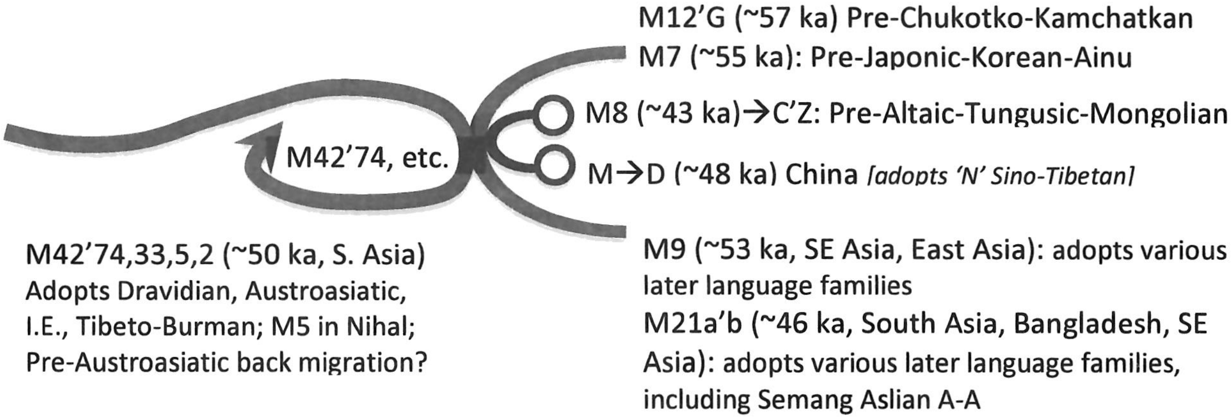
The emergence of R-mtDNA around 74 ka appears to have occurred in the Persian Gulf Oasis before the Zagros Crossing to South Asia and it underwent multiple branchings around 65 ka. A correlation for these Early and Middle Borean branches to any surviving language family does not appear evident, as it seems the early mtDNA lineages on the distal side of the Crossing in South Asia have adopted Dravidian, Indo-European or Austro-Asiatic languages. A case may be made on the 'back movement' proximal side for R2'JT (~55 ka), which seems to be associated with Pre-Semitic-Egyptian. Remarkably, the highest frequency locations of South Asian R30, R31, R5 and R6 lineages occur across northern and central India, and this suggests that rather than some sort of 'southern route' across India to E/SE Asia dispersals followed the more direct Ganges Basin and Narmada Basin routes, with outliers dispersing into Nepal and Tamil Nadu.

3. Ganges Delta Crossing to SE Asia (circa 50 ka, differentiating Middle Borean-R).


R (~54 ka, E Asia)→B4'5 (~50 ka)→B4
(~44 ka) SE Asia in Austronesian, A-A)
B4'5→B5 (~42 ka, in Kam-Tai, Tibeto-Burman)
R7,8
R→R9,21 (~47 ka), R9→F (~43 ka): Pre-Austric
R9 in Hmong-Mien, Batek, Aslian Malay, Semang;
R21 in Semang, Senoi; F in Mon-Khmer, Tai A-A
R→R7,8 (<42 ka, NE, N
India): in Vedda, Munda,
Austro-Asiatic, Dravidian
R (~58 ka, Melanesia/Australia)→P (~55 ka)
P→P4 (~53 ka, Australia): Pama-Nyungan
thus = 'southern route' via East/SE Asia

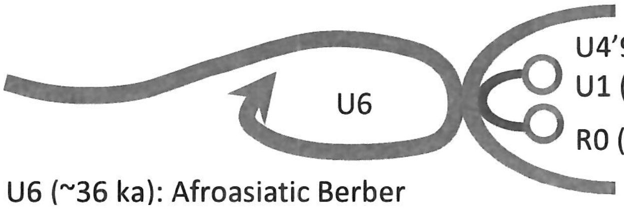
Around 50 ka R-mtDNA encountered the geographic bottleneck at the Ganges Delta Crossing to SE Asia, and this appears to correlate to a further differentiation of Middle Borean-R macrofamilies. Out of the bottleneck's distal side language families diverge into East Asia, SE Asia and Sahul. Current languages spoken by populations with high frequency and distinctive B4'5, R9 and R21-mtDNA subclades all appear to correlate to languages that have been reconstructed to the Austric macrofamily. Additionally R and its P subclades appear to arrive in the Melanesia, New Guinea and Australia around 58-53 ka and current populations distinctively P4 are currently speakers of Greater-Pama-Nyungan languages. (I discuss this in more detail under #7 Sunda/Sahul Crossing below.) On the 'back movement' proximal side of the Ganges bottleneck R7 and R8-mtDNA emerge in southern and eastern India, with high or distinctive frequency in Vedda and Munda, currently speaking diverse (perhaps adopted) languages.

4. Ganges Delta Crossing to SE/E Asia (circa 50 ka, differentiating Middle Borean-M).


M12'G (~57 ka) Pre-Chukotko-Kamchatkan
M7 (~55 ka): Pre-Japonic-Korean-Ainu
M8 (~43 ka)→C'Z: Pre-Altaic-Tungusic-Mongolian
M42'74, etc.
M→D (~48 ka) China *[adopts 'N' Sino-Tibetan]*
M42'74,33,5,2 (~50 ka, S. Asia)
Adopts Dravidian, Austroasiatic, I.E., Tibeto-Burman; M5 in Nihal; Pre-Austroasiatic back migration?
M9 (~53 ka, SE Asia, East Asia): adopts various later language families
M21a'b (~46 ka, South Asia, Bangladesh, SE Asia): adopts various later language families, including Semang Aslian A-A

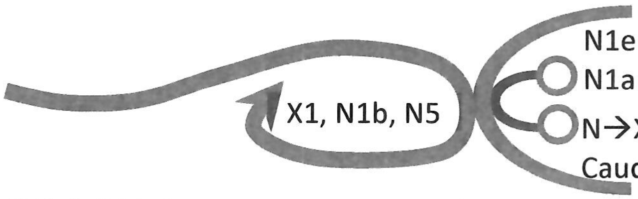
Meanwhile, around 50 ka, haplogroup M appears to have encountered the Ganges Delta bottleneck. On the distal side of the Crossing, branching southerly into SE Asia and northerly into E Asia, M subclades appear distinctively associated with several Middle Borean-M language macrofamilies. In the northerly dispersal, M subclades 12'G has high frequency and distinctiveness that correlates to Pre-Chukotko-Kamchatkan; M7 to Pre-Japonic-Korean-Ainu; and M8 to Pre-Altaic-Tungusic-Mongolian. In the southerly direction, M9 later adopts various SE and E Asian language families. M21a'b in SE Asia occurs in Semang, currently speakers of Aslian Austro-Asiatic. In East Asia D-mtDNA (frequent in Han Chinese) is a remarkably anomalous case (the exception that proves the rule?). Since this clade is currently associated with the Sino-Tibetan language family, this does not match the M subclade genetics. I hypothesize that this population group dropped an initial Eurasiatic language and adopted a Borean-N (Pre-Dené-Caucasic-Burushaski) language, which evolved into Chinese. On the proximal side of the Ganges Delta Crossing, M42'74, M33, M5 and M2 have TMRCA dates around 50 ka. They may be viewed as the 'back movement' at the geographic bottleneck, and high or distinctive frequencies of these haplogroups occur in populations that appear to have adopted various languages, including Dravidian, Austroasiatic, Indo-European and Tibeto-Burman. In part, it appears to be a Pre-Austroasiatic back migration from SE Asia. Interestingly, M5 is frequent in Nihal speakers.

5. SW Asia to W Eurasia, the Transcaucasus Crossing (circa 50 ka, differentiating Middle Borean-U; R→U ~54 ka).


U8 (~50 ka, Europe, Anatolia): Pre-Hattic
U5 (~36 ka, Europe, Russia): Pre-Finno-Ugric
U4'9 (~43 ka, C. Asia, Cauc., Europe): Pre-Kartvelian
U1 (~37 ka, SW Asia; Zagros): Pre-Elamitic (?)
U6
R0 (~39 ka, SW Asia, Iran)/HV (~27 ka): Pre-Basque
U6 (~36 ka): Afroasiatic Berber
U2 (~54 ka, Iran, S. Asia): Pre-Dravidian
U3 (~41 ka, SW Asia, Caucasus): Pre-NW-Caucasic

R's subclade U emerged in SW Asia with a TMRCA ~54 ka. Major branchings of its subclades occurred between 50 and 40 ka apparently in response to the geographic bottleneck at the Transcaucasus Crossing. On the distal side of the crossing an array of subclades dispersed easterly and westerly, with some peoples remaining in the Caucasus area. Each of these mtDNA clades has high or distinctive frequency in current populations that appear to correlate to a major language family. Based on mtDNA haplogroup frequencies in various current population samples, recent genomic analyses of fossil hominins, and archaeology, I infer—in chronological order—U2 (~54 ka) correlates to Pre-Dravidian, U8 (~50 ka) to Pre-Hattic, U4'9 (~43 ka) to Pre-Kartvelian, U3 (~41 ka) to Pre-Northwest-Caucasic, RO (~39 ka) to Pre-Basque, U1 (~37 ka) to Pre-Elamitic (?) and U5 (~36 ka) to Pre-Finno-Ugric. Again I use the designation 'Pre' to indicate a genetic TMRCA dating for a language family that may have emerged at that date or sometime thereafter. I acknowledge that linguists may view Kartvelian as emerging tens of millennia later than U4'9's date around 43 ka. On the proximal side of the Transcaucasus bottleneck I suggest U6 (~36 ka), which represents a back-migration into Africa, across Northern Africa, has archaeological correlates and in term of language may correlate to an early Afroasiatic Berber.

6. 'Silk Route Crossing', Persian and Indus Spurs then bifurcation west to Europe, east to Central and East Asia (circa 25 ka, differentiating Late Borean-N; N→N1'5 ~57 ka; N1 ~53 ka; N2 ~44 ka).


N2→N2a (>21 ka, Iran, Caucasus, E. Europe)
and W (~21 ka, W. Eurasia)
→W5 (~13 ka, Europe)
N1e (~26 ka, Russia, Buryat) and I (~26 ka, Europe);
N1a (~21 ka, SW Asia, Europe, C Asia): *adopt various*
X1, N1b, N5
N→X (~31 ka)→X2 (~21 ka, SW Asia, Europe, Caucasus, Hunza, Evenk): *adopt various languages*
N1b (~21 ka): Iran, Marsh Arab Iraq *(adopt various)*
N5 (~37 ka): Madhya Pradesh Sahariya (*adopts I.E.)*
X1'3 (~32 ka)→X1 (>25 ka, Near East, N. and E. Africa): *adopts various local languages*
N→A (~29 ka, Caspian-Baikal): Pre-Dené
N9'Y→Y (~22 ka): Nivkh (and in Ulchi, Ainu)

At about 25,000 years ago, the differentiations of N and N1-mtDNA appears to have occurred in response to northerly dispersal of N on routes later known as the Persian and Indus Spurs of the Silk Route, and, when arriving at the main Silk Route, bifurcating to disperse westerly toward and into Europe and easterly toward Central and East Asia. N2 offspring N2a and W have high frequency, if not homelands, in Eastern and Western Europe, and given current language association seem to have adopted various languages as they arrived at their current lands. In contrast A has a Caspian-Baikal homeland, and I would correlate it to Pre-Dené, and N9'Y's offspring Y has highest frequency in Nivkh. X2 as well as N1a and N1e occur in high or distinctive frequencies among current populations residing along all the east, west, north and south 'Silk' routes, and appear to have adopted their current languages. On the proximal side of the Persian and Indus Spurs geographic bottleneck, possibly to be viewed as southerly back migrations, I suggest positing the emergence of N1b with high or distinctive frequencies in Iran and Iraq Marsh Arabs, who have been proposed as descendents of the earlier Sumerians (Al-Zahery, Pala et al., 2011) and X1'3 and offspring X1, which occur predominantly in the Near East and Northern and Eastern Africa. Tentatively, I suggest that N5-mtDNA, distinctive in the Sahariya of Madhya Pradesh and considered autochthonous for India, also may represent a proximal back movement from the Indus Spur geographic bottleneck.

7. Sunda/Sahul Crossing (circa 50 ka, differentiating Middle Borean-N, M and R→P ~55 ka).

M29'Q (~44 ka, Australia/Melanesia)
→Q2: (~30 or 45 ka, Irian Jaya, New Britain):
Pre-Papuan (Kimberley Kalumburu adopt non-P-N)
→Q (~32 or 46 ka, Irian Jaya Ok-Muyu, Asmat, Una, Dani)→Q1 (~21 or 27 ka, PNG Bandi):
Trans-New-Guinea.

P→P1 (~33 or 43 ka, Irian Jaya 'pygmoid' Una, Ketengba; Bandi): *adopt TNG, Austronesian, Papuan*
P→P4 (~53 ka, Austr./Melan.)→P4b (~40/47 ka, Desert Warlpiri and NSW); P→P8 (~40 ka, Warlpiri):
Pama-Nyungan

M→M9 (~53 ka, E & SE Asia, Sunda)→M9ab'E (~47 ka): *adopt Sinitic, T-B, Tai-Kadai, Hmong-Mien, Austronesian, Papuan*
M→M21 (~46 ka, South Asia, Bangladesh, SE Asia, Semang): *adopt Austronesian, Aslian Austro-Asiatic*
M→M22 (~41 ka, Temuan Malay): *adopt Austronesian, Malayan*

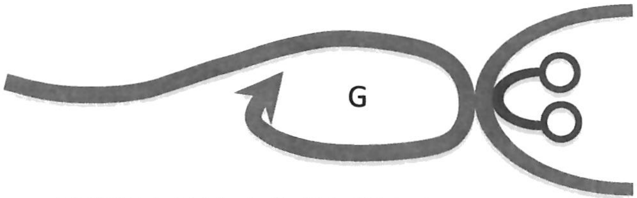
N→S (~53 ka, Australia, Tasmania):
in Pama-Nyungan and non-PN
N→O (~48 ka, SW & Desert Australia):
in Pama-Nyungan speakers

M (~53 ka, Australia/Melanesia)
→M42a (~41 ka or 33 ka, NSW Australia, Tasmania): *in Pama-Nyungan speakers*

Around 50 ka N, R and M-mtDNA encountered the geographic bottleneck at the Sunda/Sahul Crossing to New Guinea and Australia. Out of the bottleneck's distal side multiple mtDNA branches emerge as peoples disperse throughout the Sahul in northerly, centrally and southerly directions. It appears that all three major mtDNA lineages, N, R and M, enter the Sahul region. R subclade P4 appears to arrive first ~53 ka, dispersing into the Central Desert and southeasterly to the New South Wales area. P4 is strongly associated with speakers of Pama-Nyungan. Around the same time N subclades S (~53 ka) and O (~48 ka) disperse into the Australian Central Desert and southwest area and into Tasmania. In current populations O-mtDNA is frequent or distinctive in Pama-Nyungan speakers, S-mtDNA in both Pama-Nyungan and non-Pama-Nyungan speakers. Whether this suggests it was people bearing R subclade P4 that spread Pama-Nyungan and N subclades S and O later adopted it, or another alternative, I leave open. While the TMRCA for M in Australia/Melanesia has a similar date, it does not appear that a specific subclade arrives in Australia until ten thousand years later, as M42a (~41 ka). By around 20 ka, M again (Q1-mtDNA) arrives in northern Australia, apparently speakers of Trans-New-Guinea languages and later, N (N12-mtDNA) in the north, apparently speakers of Gunwinyguan languages. Circa 40 ka there were also northerly dispersals out of the Sunda/Sahul bottleneck into New Guinea and Melanesia of both M and R subclades. M29'Q offspring Q2 in Irian Jaya and New Britain is predominantly found among Papuan speakers and Q1, Trans-New-Guinea. Populations with R subclade P1 in Irian Jaya and Papua New Guinea are currently speakers of Austronesian, Papuan and Trans-New-Guinea languages, perhaps one or all are adopted. In sum, it appears there was a complex diffusion across the 'Southern Route' to Sahul, with 3 or more waves crossing the geographic bottleneck over a time spanning at least 35,000 years.

8. Siberia to the Americas, the Beringia Crossing (circa 25 ka, differentiating Late Borean-N, R and M).

N→X (~31 ka)→X1'2'3 (~29 ka, Caucasus, SW and Central Asia)→X2 (~21 ka, SW Asia to Siberia)→X2a'j (~17 ka; X2a ~13 ka, Americas, following glacial ice edge): X-Amerind (Algonquian)



M12'G (~57 ka) → G (~35 ka): Chukotko-Kamchatkan

N→A (~29 ka, Caspian-Baikal)→ A2 (~17 ka, N, C and S America): A-Amerind (Dené)

M8 (~43 ka, Altaic-Tungusic-Mongolian)→CZ→C (~28 ka, Caspian-Baikal, Yukaghir, Evenk, Tuvan)→C1 (~17 ka, esp. Central America): C-Amerind

R→B4'5 (~50 ka, SE Asia)→B4 (~44 ka, East and SE Asia)→B2 (~21 ka, western N and S America): B-Amerind
(Berezkin: 'Indo-Pacific' mythology ?)

M→M80'D→D (~48 ka, adopts 'N' language family, Sino-Tibetan)→D4→D4h (~21 ka, mostly South America by coastal route)→D4h3 (~18 ka, D4h3a ~13 ka) and D1 (~17 ka, mostly Chile, Argentina by coastal route): D-Amerind

Around 25 ka N, R and M-mtDNA encountered the geographic bottleneck at the Beringia Crossing from Siberia to the Americas. While Greenberg (Greenberg, 1960, 1987; Greenberg and Ruhlen, 2007; Ruhlen, 1994a, 1994b) group all Amerindian languages into three groups Amerind, Na-Dene and Eskimo-Aleut, the taxon Amerind has been challenged pro and con. I do not enter this particular debate as a linguist but with respect to the archaeogenetics I suggest it is possible to differentiate by the primary mtDNA haplogroups what I term X, A, C, B and D-Amerind. Those with the most evident correlation to a language family are X-mtDNA and Algonquian speakers and A-mtDNA and Dené speakers. From a mythological perspective, B-mtDNA, which in the Old World is found in SE Asia into E Asia, I suggest correlates with Berezkin's discovery of 'Indo-Pacific' mythology in the Americas, and especially South America (2010a, 2010b). In any case, out of the Beringia bottleneck's distal side mtDNA subclades and associated language families emerge to disperse occur across North America and along the coast from North to Central to South America, and thence inland. X2a-mtDNA with the Algonquian language family appears to have followed the northern North American glacial ice edge. B-Amerind and D-Amerind appear to have taken the southerly coastal route all the way to Tierra del Fuego. With respect to the scorpion diagram I place A2 and C1 in the middle since A2 (Dené) and C1 are each found across North, Central and South America, while C1 has highest frequencies in populations in Central America. On the 'back movement' proximal side of the Beringia bottleneck M12'G gives rise to G around 35 ka and G1 around 22 ka, with highest frequencies in Itelmen and Koryak, speakers of Chukotko-Kamchatkan.

Conclusions.

- A meta-pattern-analysis of the mitochondrial DNA phylotree and current distribution of language families indicates that over the last 200,000 years there are robust correspondences between mtDNA haplogroups and language macrofamilies. This study is a thought experiment, a top-down derivation of the *Homo sapiens sapiens* ('Proto-Human', 'Proto-World') language phylotree, which can be tested against bottom-up prehistoric linguistic reconstructions. It establishes a relative chronology for dating the emergence and branching of the global array of language macrofamilies. The language phylotree is crosschecked against archaeological data and fossil mtDNA studies, which support many of the correlations.

- The hypothesis of this study is that there is a rough 1:1 correspondence between the 200,000-year mtDNA phylotree and its TMRCA haplogroup dates and the emergence of language macrofamilies. A meta-pattern-analysis of the mtDNA phylotree, archaeogenetics and archaeology appears to support this hypothesis.

- The analysis provides a relative timeline for the emergence and branching of all the language macrofamilies of *Homo sapiens sapiens* language ('Proto-Human', 'Proto-World'), which may prove useful for linguistic reconstructions of proto-Sapiens-Sapiens and for reconstructions of the prehistory of mythological and ritual systems both within-Africa and out-of-Africa.

- Proto-Sapiens-Sapiens appears to have emerged with the earliest stage of fossil *Homo sapiens sapiens* at Early Middle Stone Age Omo Kibish, Ethiopia, around 195,000 years ago. Late dating of 'human' language origins to 45, 60 or even 100 ka is contradicted by mtDNA archaeogenetics as well as archaeology.

- In this proposed timeline, click languages, strongly associated with L0-mtDNA, diverged from all other languages around 160,000 years ago. The Niger-Congo language family, robustly correlated to L1-mtDNA, emerged around 140,000 years ago. Around 120,000 years ago the ancestor of Central Sudanic, apparently correlating to L5-mtDNA, diverged from a pre-Nilo-Saharan-Afroasiatic macrolanguage. At this time period the First Wave Dispersal out-of-Africa occurs bearing L2'3'4'6-mtDNA and Pre-Nilo-Saharan-Afroasiatic.

- Around 100,000 years ago during the period of the Lake Paleo-Chad and central Sahara corridor the ancestral divergence occurred between Afroasiatic and Nilo-Saharan macrolanguage families, the former associated with L2-mtDNA and the latter, L3'4'6. Archaeology at this stage evidences a Second Wave Dispersal out-of-Africa into SW Asia at Aybut Auwal, Oman, which would have carried one or both of these ancestral languages, proto-Austroasiatic and/or proto-Nilo-Saharan.

- Archaeology in SW Asia around 85 ka provides evidence for at least two out-of-Africa industries, including two sites in the Sinai, with Nubian Complex and Nile Denticulate Mousterian, and one in northern Saudi Arabia at the Jubbah paleolake, with Nubian Complex affinity, perhaps correlating to L3'4. This appears to be a continuation of the Second Wave dispersals out of Africa. The Nubian Complex MSA industry could have been bearers of an ancestor of the Nilo-Saharan proto-Northern Sudanic (Kunama) or proto-Koman (Gumuz, Uduk) language families (in Ehret 2011 terms) and the Nile Denticulate Mousterian, an ancestor of the Afroasiatic Boreafrasian language family.

- Around 80,000 years ago (late MIS 5a) a Third Wave Dispersal out-of-Africa occurred bearing the L3 subclades M and N. Archaeology and mtDNA genetics again imply at least Nubian Complex and Nile Denticulate Mousterian cultural traditions diffused into SW Asia, correlatable to Northern Sudanic and/or Koman and Boreafrasian language families, and also to Proto-Saharan-Sahelian (Kanuri) or Proto-Eastern Sahelian (Nubian), and these interacted with SW Asian indigenous populations having Tabun C industries. In inferring this I am not equating M and N respectively to these two traditions, but dispersing populations probably had varying admixtures of northeastern African L3M and L3N as well as northern African L2a. A similar argument would apply to their ritual-myth-and-art traditions.

- There is currently insufficient archaeological evidence to determine whether routes out-of-Africa were via the Sinai, the Bab-el-Mandeb or circum-Red Sea or some combination thereof.

- In terms of mythostratigraphy out-of-Africa, based on archaeology and mtDNA genetics, I suggest the best inference would be that the dispersals out of Africa would have carried myth-ritual systems that combined to greater or lesser extent components of the two major North and East African myth-systems, which evolved prior to the TMRCA of L3 and which may be termed 'North African' (correlating to L2, L2a) and 'Sudanic' (correlating to L5, L4) or 'Nile-Sudanic' (correlating to L3 and its subclades). The Nile-Sudanic myth-system would have emphasized ancestral and game spirits, where 'spirit' means a life-giving, life-animating and life-enhancing forces or energies. This is a religious system having neither a high god nor *deus otiosus*. The North African religious system would appear to have had a creative power, female or androgynous Creatrix, who organized the cosmos with a world-axis, four directions and their associated complementarity principles, and thereby established the nature of life's unfolding as one of balance and complementarity of polar or gendered pairs in all their variant combinations and recombinations.

- Around 75,000 years ago L3-mtDNA's out-of-Africa dispersal into SW Asia had a pause of up to 5,000 years, during which N and M differentiated subclades and N branched off R-mtDNA and its subclades. Geographic bottlenecks at the Transcaucasus Crossing to Western Eurasia and Zagros Crossing to South Asia, extant Neanderthals in both directions and other archaic species, and possibly the Toba supereruption (~74 ka) and positive subsistence landscape of the 'Persian Gulf Oasis' presumably contributed to this extended delay.

- Three major Borean language families appear robustly associated with the differentiation of N, R and M-mtDNA clades, and I term these Borean-M, Borean-N and Borean-R. Respectively, M-mtDNA corresponds strongly to the Eurasiatic language family; N-mtDNA to the Dené-Caucasian, and R-mtDNA to Afroasiatic, Kartvelian, Dravidian, etc. I suggest that Austric and Pama-Nyungan are most closely associated with R-mtDNA. Borean-R appears to have taken a 'northern route' diffusing branches across northern South Asia and into Southeast Asia and Sahul, and also into Europe, Central and East Asia. While at first glance this might appear contradictory to current language macrofamily reconstructions, it seems supported by Berezkin's discovery using principle component analysis (2010a, 2010b) of similar mythological motifs shared by peoples in three disparate geographic regions: Indo-Pacific, South American and western Eurasia/Europe.

- The hypothesis for a *sapiens sapiens* 'southern route fast track' of a few thousand years from SW Asia to Sahul, e.g., leaving Africa around 50, 60 or even 70 ka is not supported by mtDNA genetics or archaeology. On the contrary the Third Wave Dispersal from

Africa began circa 80 ka and arriving in Sahul circa 55 ka apparently took around 25,000 years, including an up to 5,000 year delay in SW Asia.

- By around 50,000 years ago (MIS 3c/b), based on mtDNA phylotree and its TMRCAs, it appears that roughly 28 new mtDNA haplogroups had emerged, and based on current languages associated with them, at least 17 of them, from Europe to East Asia and Sahul, correlate to the emergence of 17 language families—which I term 'Middle Borean'.

- By around 25,000 years ago another dozen language families emerged—which I term 'Late Borean'—and this, for the most part, appears to have completed the development of the major language families of the world.

- The 1:1 correlation of mtDNA haplogroups and language families generally appears more robust at earlier stages of the phylotree than recent stages. In the latter language replacements and adoptions appear more frequent and current population genetic samples sometimes more admixed. There is one notable anomaly to the 1:1 correlations for Middle Borean languages around 50,000 years ago. Current Han people are especially associated with mtDNA haplogroup D, which belongs to the M-mtDNA clade. They would be expected to have a Borean-M Eurasiatic language. Instead their Chinese language is classified as belonging to the Dené-Caucasian family (Borean-N). Thus the genetics suggests that the Han may have initially been speakers of a Eurasiatic language and later adopted a Dené-Caucasian language, which evolved into Chinese. Linguists might explore this possibility further. While the case of Han Chinese might be taken to invalidate my basic hypothesis of a 1:1 correlation of major mtDNA haplogroups and language macrofamilies, I suggest that the correlation appears to hold in general and this Han Chinese exception appears to be the exception that proves the rule.

- Finally, it appears that Fleming's Borean model (Fleming, 2002; 1991; 1987; Fleming, Zegura et al., 2013) with its 3 major subclades maps almost precisely onto the 3 primary branches of mtDNA out-of-Africa and their correlated language macrofamilies. Fleming's cluster of Afrasian (Afroasiatic), Kartvelian, Dravidian, Elamitic, and other SW Asia extinct languages maps onto descendents of R-mtDNA and more precisely its U-mtDNA branches; Caucasic-Burushaski-Dené maps onto N-mtDNA and Eurasiatic onto M-mtDNA. Archaeogenetics further supports adding SE Asian/Sahul languages, including Austric, Trans-New-Guinea, Papuan and Pama-Nyungan, as a fourth cluster to Fleming's Borean, as argued for by Gell-Mann, Peiros and Starostin (2009) and correlating these languages to the Borean-R language phylum.

Limitations.

Evidence for correlating some languages to mtDNA haplogroups appears more robust than others; it seems by and large that correlations become weaker the closer we approach recent times. As a nonlinguist, I am aware that there are complex language replacement issues pertaining to mismatches between the languages populations currently speak and those they may have spoken in historic or prehistoric times. The language correlations that I have proposed are more or less tentative. I am open to counterarguments with respect to which language to correlate to particular branches in the mtDNA phylotree.

In this study I have chosen to first look at the archaeogenetics of mtDNA—or so to speak, the 'mother tongue'. The archaeogenetics of Y-DNA remains for future research.

Because of the complexity of cross-mapping databases for mtDNA archaeogenetics, language macrofamilies, and archaeology and the constant advances in each field, the meta-mapping proposed in this study is necessarily tentative and open to revision.

Acknowledgments.

I offer my gratitude to Hal Fleming, and colleagues of the *Mother Tongue* journal, for the challenge and encouragement to bring my interest in mythology, archaeology and archaeogenetics to the question of long-range language origins. I thank Luca Pagani for generating the haplogroup assignments and frequencies for his Gumuz sample that I might refer to them in this study. Also thanks to the other geneticists who answered over the years emails in which I asked so many neophyte questions. May this study also honor a mentor, the archaeologist, linguist and folklorist Marija Gimbutas, who inspired and encouraged me to enter the field of paleolithic myth and symbol.

Supplementary Information.

The mtDNA Database for Archaeogeneticlinguistics containing approximately 181 pages, 424 ethnic and population mtDNA haplogroup frequency samples, their current spoken languages, 82 fossil mtDNA studies, and select archaeological sites, with references, is available online at https://originsnet.academia.edu/JamesHarrod)

APPENDIX 1: Out-of-Africa mtDNA and Correlated Language Family Dispersal Maps

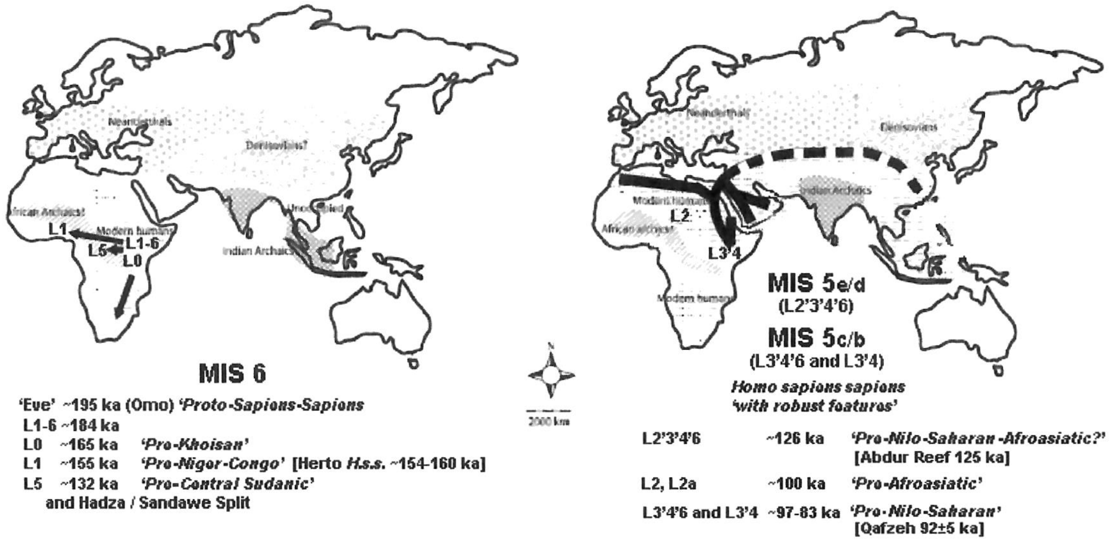

Neanderthals
Denisovans?
African Archaics?
L1
Modern humans
L1-6
L5
L0
Unoccupied
Indian Archaics
MIS 6
'Eve' ~195 ka (Omo) *'Proto-Sapiens-Sapiens*
L1-6 ~184 ka
L0 ~165 ka *'Pro-Khoisan'*
L1 ~155 ka *'Pro-Niger-Congo'* [Herto *H.s.s.* ~154-160 ka]
L5 ~132 ka *'Pro-Central Sudanic'*
and Hadza / Sandawe Split
N
2000 km
Neanderthals
Denisovans
Indian Archaics
Modern Humans
L2
African archaics
L3'4
Modern humans
MIS 5e/d
(L2'3'4'6)
MIS 5c/b
(L3'4'6 and L3'4)
Homo sapiens sapiens
'with robust features'
L2'3'4'6 ~126 ka *'Pro-Nilo-Saharan -Afroasiatic?'* [Abdur Reef 125 ka]
L2, L2a ~100 ka *'Pro-Afroasiatic'*
L3'4'6 and L3'4 ~97-83 ka *'Pro-Nilo-Saharan'* [Qafzeh 92±5 ka]

Figure 1. MIS 6 (~190-130 ka). Early *Homo sapiens sapiens* in East Africa and other archaics in Africa. Neanderthals in Europe and Denisovans in Central and Eastern Asia, Indian archaics in the Indian Subcontinent and Sundaland.

MIS 5e/d (~130-106 ka) and MIS 5c/b (~106-85 ka). Based on archaeogenetics and archaeology *Homo sapiens sapiens* 'with robust/archaic features' and Middle Paleolithic technologies expand across Northern Africa and out-of-Africa into SW Asia in what appears to be two successive waves. L2'3'4'6 diffuses across Northern Africa (Magreb Levallois Mousterian and Aterian; Bir Tarfawi Aterian and Early Nubian Complex; Abdur MSA with handaxes) and into SW Asia (Jebel Faya with handaxes). Fossil Zhirendong, south China, at minimum ~106 ka, suggests expansion across South Asia into East and SE Asia. A second diffusion, probably L3'4'6, into SW Asia occurs during MIS 5c/b (Nazlet Khater Nubian Complex; Sinai Split Rock-Lower Denticulate Mousterian; Aybut Auwal Nubian Complex). Indian archaics retreated from Sundaland submerged by higher sea level, Mishra, Chauhan and Singhvi (2013).

mtDNA haplogroup TMRCAs from Soares, Ermini, et al (2009), with caveat that for MIS 6 and MIS 5e/d Soares TMRCAs are multiplied by a factor of 1.1x to accomodate redating of chimpanzee/human split (Langergraber, Prüfer et al 2012; their redating is 1.2x CHCLA used by Soares; to be conservative I use a 1.1x multiplier). Also for L3'4'6 and L3'4 I have averaged Soares TMRCAs with Behar, van Oven, et al (2012) TMRCAs times 1.1. Correlations of haplogroups to ancestors of associated language families are author's tentative hypotheses. Maps are author's archaeogenetic overlays and modifications of maps in Mishra, Chauhan and Singhvi (2013: fig.2), which provide a summary display of global population movements with respect to South Asia and challenge the single-diffusion southern route fast track out-of-Africa to SE Asia model.

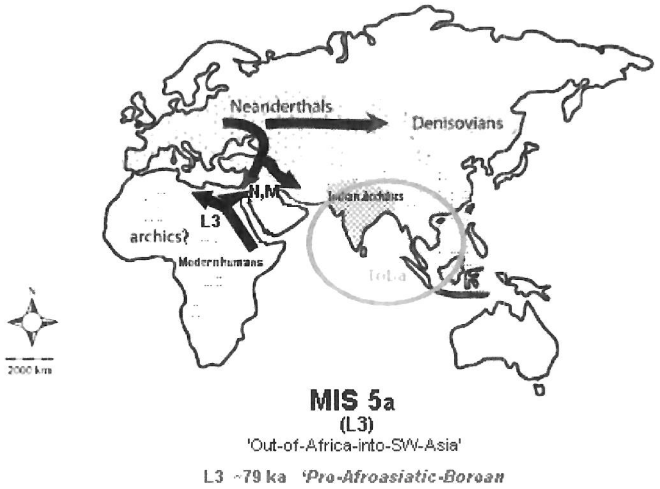
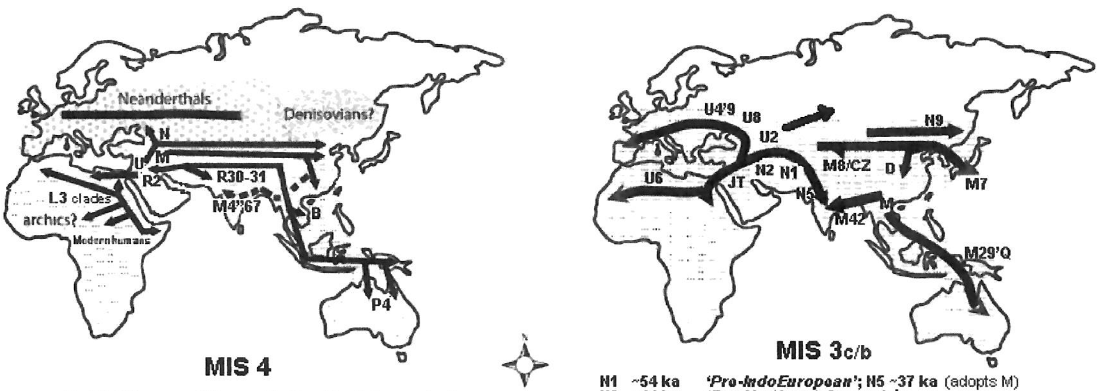

Neanderthals
Denisovians
N,M
Indian Archaics
L3
archics?
Modern humans
Toba
N
2000 km
MIS 5a
(L3)
'Out-of-Africa-into-SW-Asia'
L3 ~79 ka 'Pro-Afroasiatic-Boroan

Figure 2. MIS 5a (~85-74 ka). Around 80 ka a third wave of *Homo sapiens sapiens* with haplogroup L3 and Middle Paleolithic technologies diffuses across Northern Africa (Ifri n'Ammar Aterian; Pigeons Taforalt Aterian; El-Guettar 'Final Mousterian'; Bir Tarfawi Aterian; Taramsa 1-Phase III Levallois and Nubian) and out-of-Africa into SW Asia (possibly Jebel Qattar Tabun C; Sinai Split Rock Upper Denticulate). Bottlenecks due to geography, Neanderthal movement into SW Asia and the Toba supereruption (~74 ka) prevented further diffusion. The latter also contributed to fading away of the Indian archaics.

L3 TMRCA from Soares, Ermini, et al (2009) is 72 ka, but Soares, Alshamali, et al (2012) with respect to 'expansion of subclades of L3' is 65 (or 66 for earliest L3h). Behar, van Oven, et al (2012) calculates a series of clock violations around L3 out-of-Africa dating, with L'3'4'6 71 ka, L3'4 64 ka, L4 79 ka and L3 67 ka. Given the range of dates for L4 and L3, if the outliers are eliminated the average of L3 dates at 72 and 67 x 1.1 = 76 ka. I have rather taken Soares, Ermini, et al (2009) L3 at 72 x 1.1 = 79 ka.



N ~75-78 ka	***'Boroan-N: Pro-Dené-Caucasian'***
M ~67 ka	***'Boroan-M: Pro-Eurasiatic'***
N>R ~74 ka	***'Boroan-R'***
R31-30 ~71 ka	
U ~59 ka	***'Pro-Dravidian/Kartvolian'***
R2'JT ~60 ka	
B ~56 ka	***'Pro-Austric'***
P ~64 ka	***'Pro-Pama-Nyungan'***
M4"67 ~62 ka; M2 ~60 ka	

N1 ~54 ka	***'Pro-IndoEuropean'*; N5 ~37 ka** (adopts M)
N2 ~44 ka	***'Pro-Northeast Caucasic'***
N9 ~49 ka	***'Pro-Nivkh (Udogay)'*** (adopts M-macrolanguage)
M7 ~55 ka	***'Pro-Japonic-Korean-Ainu'***
D ~48 ka	***'Pro-Sino-Tibetan'*** (D adopts N-macrolanguage)
M8/CZ ~43 ka	***'Pro-Altaic-Tungusic-Mongolian'***
M42'74 ~55 ka	***'Pro-Austro-Asiatic'*; M33 ~45 ka**
M in SE Asia ~41-52	***'Austric'***
M29'Q ~47	***'Pro-Trans-New-Guinea'***
U2 ~55 ka	***'Pro-Dravidian'*** (Eastern Aurignacian ~40 ka)
U8 ~50 ka	***'Pro-Hattic'***
U4'9 ~45 ka	***'Pro-Kartvolian'***
JT ~50 ka	***'Pro-Semitic/Egyptian'***

Figure 3. MIS 4 (~74-59 ka) and MIS 3c/b (~59-40 ka). During MIS 4, L3 subclades expand in Africa, while in SW Asia, after about a 4,000 year stasis and post-Toba, modern humans disperse out-of-SW-Asia in at least three waves along three northern routes. M and N diverge as they migrate mostly north of the Himalayas along the 'silk route'. Shortly thereafter remaining N in SW Asia branches off R which further branches off U clades in SW Asia and moves south of the Himalayas across northern India, branching off R31-30, and from there into SE Asia and eventually Australia. Presumably all three of these MIS 4 waves carry Middle Paleolithic tool-and-art kits. Archaic populations contribute some admixture into these dispersing modern human lineages.

During MIS 3c/b the three out-of-SW Asia M, N and R lineages continue branching and migrating, with significant back migrations in Africa and SE to South Asia. During MIS 3c/b archaic populations disappear with some admixture into modern humans. By the end of MIS 3c/b, I suggest, most of the Eurasian language families have evolved.

TMRCAs for MIS 4 for the most part are from Soares, Ermini, et al (2009) conservatively multiplied by 1.1 for the revised CHCLA, with several caveats. Soares, Ermini, et al (2009) dates N at 71 x 1.1 = 78 ka; but Soares, Alshamali, et al (2012) reduces this to 62 x 1.1 = 68 ka. This reduction results in a clock violation with the earlier Soares TMRCAs for R-mtDNA (67 x 1.1 =74 ka) and all subsequent branches. Thus, I suggest the range 75-78 ka between R at 74 ka and L3 at 79 ka. For M4"67 and M2 in South Asia, rejecting the low dating of both Soares (2009) and Behar (2012), I select the TMRCAs from Thangaraj, Chaubey, et al (2006; compare Kumar, Padmanabham, et al 2008).

TMRCAs for MIS 3c/b are all taken from Soares, Ermini, et al (2009) and I have applied no multiplier, as TMRCAs during this more recent time period seem satisfactory with respect to archaeology and use of multiplier appears to result in numerous clock violations. For N2 Soares provides no TMRCA; I use Fernandes, Alshamali, et al (2012). Similarly, for N5 and M29'Q I use Behar, van Oven, et al (2012); for M42'74 I use Kumar, Ravuri, et al (2009); and for M in Southeast Asia I use Jinam, Hong, et al (2012).

From the beginning of the next stage, MIS 3a ~40 ka, Later Stone Age/Upper Paleolithic microblade technologies occur across Africa, SW Asia, South Asia, Europe and beyond. Each region likely saw innovations by a mix of indigenous groups and cross-region diffusion. For example South Asia microblade industries may have been innovated by R31-30 and M4"67-mtDNA descendents out of their Mode 3 technologies along with diffusions via N2/N5 and U2 lineages from the northwest.

Derivation of Dates for mtDNA and Correlated Language Family Dispersal Maps						
Approximate TMRCA			**Sources and Calculations**			
MIS 6 (190-130)			**Soares**	**Gonder**	**BvO**	**Other**
Eve	195		**192** *x1.1=211*	194	177 *x1.1=**195***	[Omo-Kibish = 195]
L1-6	184		167 [L(3-7)1] *x1.1=**184***		153 [L1-5]	
L0	165		150 *x1.1=**165***	146	136	
L1	155		141 *x1.1=**155***		128.5	**[Herto = 154±7-160±2]**
L5	132		120 *x1.1=**132***	129	111	BL: 106
MIS 5e/d (130-106)						
L2346	126		115 [L34762] *x1.1=***126**		111 *x1.1=122*	**[Skhul = 100-130]**
MIS 5c/b (106-85)					L0a’b’f: 100x1.1=110	
L3’4’6	97		105 [L3476] *x1.1=115*		71 *x1.1=78*	ave 115+78=**97** **[Qafzeh = 92±5]**
L3’4	83		86 [L347] *x1.1=95*		64[L34] *x1.1=70*	ave 95+70=**83**
MIS 5a (85-74)					~~L4: 79 violation~~	
L3	79		72 (S2009) x1.1=79	~~Gonder~~ ~~97~~ ~~OOA 94~~	L3: 67 violation x1.1=74	BL: 101
			~~‘expansion’ (S12)~~ ~~65 (but L3h=66)~~		*Or if Gonder excluded, all 4 dates L3 and L4 (65, 67, 72 and 79 ave = 71 x1.1=78; or if elim. outliers then ave 67+72=69 x1.1=76*	
MIS 4 (74-59)			*[multiplier x1.1]*			
N	75-78		71SoAs x1.1=**78** ~~S1262 x1.2=74 / x1.1=68~~ 58EA [S12 yields clock violation on R; ergo N = betw 74 and 79]		59	
M	67		60.6 x1.1=**67** S1260.5x1.2=73 / x1.1=67 49SA		50	J: 64
N→R	74		67SoA x1.2=80 / 1.1=**74**		56.5 geoi indet x1.2=68 / x1.1=62	
R31 **R30**	71		64.5 64 NNWSA x1.2=77/1.1=**71**		R31: 55 R30: 54 55x1.2=66 / x1.1=61	
U	59		54 x1.2=65 x1.1=**59**		47 x1.2=66 1.1=52	A: 59
R2JT	60		55 x1.2=66 / x1.1=**60**		54 x1.2=65 1.1=59	
B	56		B: 51x1.2=61/x1.1=**56** R=54EA		B4’5: 49.5 x1.2=59 / 1.1=54	
P	64		P: 58Aus/Mel x1.2=70/1.1=**64**		P: 55, P4: 53	H: P=52 / P4=66!
M4”67	62		40.5 x1.2=49 / x1.1=44 ?49SA x1.2=59x / 1.1=54		35 *x1.2=42/ 1.1=38*	**T: M4’30: 62**
M2	60		38 x1.2=46 / x1.1=42		36 *x1.2=43/1.1=40*	**T: 60** KP: 50
MIS 3c/b (59-40)			*[no multiplier]*			
N1	54		**54**		N1’5: 57	FA: 50-63
N2	44-51				44 N5: **37**	FA: **44-51**
N9	49		**49**		46	SV: 39
M7	55		**55**		45	
D	48		**48**		38	
M8/CZ	43		**43**		36	
M42’74	55				50	KR: **55**
M33	45		**45**		42	T: 57
M-SEA	41-52				H: 53 Aust/Mel; J: **41-52**	
M29’Q	47				**47**	H: 44
U2	55		**55**		43	
U8	50		**50**		43	
U4’9	45		**45**		37	
JT	50		**50**		47	P: 56-58

OLD L3 calculation from MT 2013
If all 4 dates L3 and L4 ave = 71x1.2= 85; if elim outliers, 67 and 72 ave=69.5x1.2 = ***83****;*
*if 65x1.2=****78*** *or 66x1.2=79 or 72x1.2=86*

[A] Achilli, Rengo, Magri et al., 'H Franco-Cantabrian Refuge' 2004
[BvO] Behar, van Oven et al., 'Copernican Reassessment', 2012.
[BL] Batini, Lopes et al., 2010 (2011).
[FA] Fernandes, Alshamali et al., 2012
[G] Gonder, Mortensen et al., 2007. Table 2. Uses Sanderson 1997, 2002, 2003; and CHLCA 6.0+0.5 MYA.
[H] Hudjashov, Kivisild et al., 2007 using Kivisild 2006 dating
[J] Jinam, Hong, Phipps et al., 2012
[KP] Kumar, Padmanabham, Ravuri et al., 2008
[KR] Kumar, Ravuri, Koneru et al., 2009
[P] Pala, Olivieri, Achilli, et al., 2012
[S12] Soares, Alshamali et al., 2012
[S] Soares, Ermini et al., 2009. Uses complete genome clock; average of methods, 1 mutation per 7884 years; CHLCA 6.5+0.5 =7 MYA. I have taken Point Estimate dates from Sup. Info. Table S5, which uses complete genome and maximum likelihood analysis; and also indicate Table 3 dates for L3, M, N, R and U, which use overlapping interval (OvInt) of complete genome rho from two different clocks and maximum likelihood analysis estimates.
[SV] Sukernik, Volodko, Mazuin et al., 2012
[T] Thangaraj, Chaubey, Singh et al., 2006

References

(Only references directly mentioned in this narrative; references for the Supplementary File, Table 1 Master Database table run into the hundreds and can be consulted in that table, http://www.originsnet.org/publications.html.)

Abu-Amero, K. K., Larruga, J. M., Cabrera, V. M., & González, A. A. (2008). Mitochondrial DNA structure in the Arabian Peninsula. *BMC Evolutionary Biology*, *8*(45), e1-15.

Achilli, A., Perego, U. A., Bravi, C. M., Coble, M. D., Kong, Q. P., Woodward, S. R., ... Bandelt, H. J. (2008). The phylogeny of the four pan-American MtDNA haplogroups: implications for evolutionary and disease studies. *PLoS One, 3*(3), e1764.

Achilli, A., Rengo, C., Magri, C., Battaglia, V., Olivieri, A., Scozzari, R., ... Torroni, A. (2004). The molecular dissection of mtDNA haplogroup H confirms that the Franco-Cantabrian glacial refuge was a major source for the European gene pool. *American Journal of Human Genetics, 75*(5), 910-918.

Aouadi-Abdeljaouad, N., & Belhouchet, L. (2006). Nouvelles recherches préhistoriques en Tunisie Centrale: les occupations préhistoriques dans la région de Maknassy. Session C78-06. UISPP XV Congress. Lisbon.

Amos, W., & Hoffman, J. I. (2009). Evidence that two main bottleneck events shaped modern human genetic diversity. *Proceedings of the Royal Society B: Biological Sciences*, rspb20091473.

Anati, E. (2006) *Har Karkom Survey Report. Har Karkom: 20 years of biblical archaeology*. Valcamonica (BS), Italy: Edizioni del Centro. http://www.harkarkom.com/

Armitage, S. J., Jasim, S. A., Marks, A. E., Parker, A. G., Usik, V. I., & Uerpmann, H-P. (2011). The Southern Route "Out of Africa": Evidence for an Early Expansion of Modern Humans into Arabia. *Science, 334*(488), 453-456.

Atkinson, Q. D. (2010). The prospects for tracing deep language ancestry. Interdisciplinary views on Molecular Anthropology in the Genomic Era. *Journal of Anthropological Sciences, 88*, 231-233

Atkinson, Q. D. (2011). Phonemic Diversity Supports a Serial Founder Effect Model of Language Expansion from Africa. *Science, 332*(6027), 346-349.

Avery, D. M. (2003). Early and Middle Pleistocene environments and hominid biogeography; micromammalian evidence from Kabwe, Twin Rivers and Mumbwa Caves in central Zambia. *Palaeogeography, Palaeoclimatology, Palaeoecology, 189*(1), 55-69.

Bada, J. L., Schroeder, R. A., Protsch, R., and Berger, R., (1974). Concordance of Collagen-Based Radiocarbon and Aspartic-Acid Racemization Ages. *Proceedings of the National Academy of Sciences USA, 71*, 914–917.

Barham L. (1995). The Mumbwa Caves Project, Zambia, 1993-94. *Nyame Akuma, 43*, 66-72.

Barham, L. (2000). *The Middle Stone Age of Zambia, South Central Africa*. Bristol: Western Academic and Specialist Press.

Barham L. (2002a). Systematic pigment use in the Middle Pleistocene of South-Central Africa. *Current Anthropology, 43*(1), 181-190.

Barham L. (2002b). Backed tools in Middle Pleistocene central Africa and their evolutionary significance. *Journal of Human Evolution, 43*(5), 585-603.

Barham, L. (2012). Clarifying Some Fundamental Errors in Herries' "A Chronological Perspective on the Acheulian and Its Transition to the Middle Stone Age in Southern Africa: The Question of the Fauresmith" (2011). *International journal of evolutionary* biology, Article ID 230156, 5 pages.

Barham, L., & Debenham, D. (2000). Mumbwa Caves chronology. In L. Barham (ed.) *The Middle Stone Age of Zambia, South Central Africa* (pp. 43-49). Bristol: Western Academic and Specialist Press.

Barham, L., Pinto Llona, A.C., & Andrews, P. (2000). The Mumbwa Caves Behavioural Record. In L. Barham (ed.) *The Middle Stone Age of Zambia, South Central Africa* (pp. 81-148). Bristol: Western Academic and Specialist Press.

Barham, L.S., Pinto Llona, A.C., &Stringer, C.B. (2002). Bone tools from Broken Hill (Kabwe) cave, Zambia, and their evolutionary significance. *Before Farming, 2002*(2,3), 1-12.

Barton, R.N.E., Bouzouggar, A., Collcutt, S.N., Schwenninger, J-L., & Clark-Balzan, L. (2009). OSL dating of the Aterian levels at Dar es-Soltan I (Rabat, Morocco) and implications for the dispersal of modern *Homo sapiens*. *Quaternary Science Reviews, 28*, 1914–1931.

Bar-Yosef, O., & Vandermeersch, B. (1993). Modern humans in the Levant. *Scientific American, 268*(4), 94.

Batini, C., Lopes, J., Behar, D.M., Calafell, F., Jorde, L.B., van der Veen, L., … Comas, D. (2011). Insights into the demographic history of African Pygmies from complete mitochondrial genomes. *Molecular Biology and Evolution, 28*(2), 1099-110.

Behar, D. M., Harmant, C., Manry, J., van Oven, M., Haak, W., Martinez-Cruz, B., ... & Quintana-Murci, L. (2012). The Basque paradigm: genetic evidence of a maternal continuity in the Franco-Cantabrian region since pre-Neolithic times. *The American Journal of Human Genetics, 90*(3), 486-493.

Behar, D. M., van Oven, M., Rosset, S., Metspalu, M., Loogväli, E.L., Silva, N.M., … Villems, R. (2012). A "Copernican" reassessment of the human mitochondrial DNA tree from its root. *American Journal of Human Genetics, 90*(4), 675-84

Behar, D. M, Villems, R, Soodyall, H., Blue-Smith, J, Pereira, L., Metspalu, E., … Rosset S.; Genographic Consortium. (2008). The dawn of human matrilineal diversity. *American Journal of Human Genetics, 82*(5), 1130-40.

Belfer-Cohen, A., & Hovers, E. (1992). In the eye of the beholder: Mousterian and Natufian burials in the Levant. *Current Anthropology*, 463-471.

Bender, L. M. (1997). The Nilo-Saharan Languages: A Comparative Essay. Munich: Lincom Europa.

Bender, L. M. (2000). Nilo-Saharan. In B. Heine & D. Nurse (eds.), *A Linguistic Geography of Africa* (pp. 43-73). Cambridge: Cambridge University Press.

Benedict, P. K. (1976). Austro-Thai and Austroasiatic. In Philip N. Jenner, Laurence C. Thompson & Stanley Starosta (eds.), *Austroasiatic Studies, Part I* (pp. 1–36). Honolulu: University of Hawai‘i Press.

Bengtson, J. D. (1997). The riddle of Sumerian: A Dene–Caucasic language? *Mother Tongue, 3*, 63-74.

Bengtson, J. D. (2008). Materials for a Comparative Grammar of the Dene–Caucasian (Sino-Caucasian) Languages. *Aspects of Comparative Linguistics (Moscow), 3*, 45–118.

Bengtson, J. D. (2010). The Greater Austric Hypothesis. In Bengtson JD, *Linguistic Fossils: Studies in Historical Linguistics and Paleolinguistics*. Calgary: Octavia Press.

Bengtson, J. D, & Ruhlen, M. (1994). Global Etymologies. In Merritt Ruhlen, *On the Origin of Languages: Studies in Linguistic Taxonomy* (pp. 277–336). Stanford: Stanford University Press.

Benito-Calvo, A., Barfod, D. N., McHenry, L. J., & de la Torre, I. (2014). The geology and chronology of the Acheulean deposits in the Mieso area (East-Central Ethiopia). *Journal of Human Evolution, 76*, 26-38.

Berezkin, Y. (2010a). From Africa and back: some areal patterns of mythological motifs. *Mother Tongue, XV*, 1-67.

Berezkin, Y. (2010b). The emergence of the first people from the underworld: another cosmogonic myth of a possible African origin. In Wim van Binsbergen & Eric Venbrux (eds), *New Perspectives on Myth: Proceedings of the Second Annual Conference of the International Association for Comparative Mythology* (pp. 109-125). Haarlem, The Netherlands: PIP TraCS, Shikanda.

Bird, M. I., Turney, C. S. M., Fifield, L. K., Jones, R., Ayliffe, L. K., Palmer, A., Cresswell, R., & Robertson, S. (2002). Radiocarbon analysis of the early archaeological site of Nauwalabila I, Arnhem Land, Australia: implications for sample suitability and stratigraphic integrity. *Quaternary Science Reviews 21*: 1061–1075.

Blench, R. M. (1995). Is Niger-Congo simply a branch of Nilo-Saharan? In Robert Nicolaï and Franz Rottland (eds.), *Proceedings of the Fifth Nilo-Saharan Linguistics Colloquium, Nice, 24-29*

August, 1992 (= Nilo-Saharan Linguistic Analyses and Documentation 10) (pp. 83-130). Köln: Köppe.

Blench, R. (2014). Combining linguistics, archaeology and genetics to reconstruct prehistory; examples from Africa and the New World, London 30 May, 2014, Sparta History Group. (accessed 9 November 2014)

Blinkhorn, J., Achyuthan, H., Petraglia, M., & Ditchfield, P. (2013). Middle Palaeolithic occupation in the Thar Desert during the Upper Pleistocene: the signature of a modern human exit out of Africa?. *Quaternary Science Reviews*, *77*, 233-238.

Blumenschine, R. J., Brandt, S. A., & Clark J. D. (1983). Excavations and analysis of Middle Palaeolithic artifacts from Patpara, Madhya Pradesh. In Sharma, G. R. & Clark, J.D (eds.), *Palaeoenvironments and Prehistory in the Middle Son Valley*. Allahabad: Abinash Prakashan, pp. 39–99.

Boattini, A., Castri, L., Sarno, S., Useli, A., Cioffi, M., Sazzini, M., … Luiselli, D. (2013). mtDNA variation in East Africa unravels the history of Afro-Asiatic Groups. *American Journal of Physical Anthropology, 150*(3), 375-385.

Bodner, M., Perego, U. A., Huber, G., Fendt, L., Röck, A. W., Zimmermann, B., Olivieri, A., ... Parson, W. (2012). Rapid coastal spread of First Americans: Novel insights from South America's Southern Cone mitochondrial genomes. *Genome Research, 22*(5), 811-820.

Bomhard, A. R. (2008). *Reconstructing Proto-Nostratic: Comparative Phonology, Morphology, and Vocabulary*, 2 volumes. Leiden: Brill.

Bouzouggar, A., & Barton, R. N. E. (2012). The Identity and Timing of the Aterian in Morocco. Chapter 7, *Modern Origins: A North African Perspective* and *Vertebrate Paleobiology and Paleoanthropology*, Part 2, 93-105.

Bouzouggar, A., Barton, N., Vanhaeren, M., d'Errico, F., Collcutt, S., Higham, T., … Starnbouli, A. (2007). 82,000-year-old shell beads from North African and implications for the origins of modern human behavior. *Proceedings of the National Academy of Sciences USA, 104*(24), 9964-9969.

Brooks, A. S. (2005). *Background to "Out of Africa 3": behavior and environmental change in the Middle Stone Age. Presentation.* Paleoclimates and Human Evolution: a workshop on integrating continental drilling research with paleoanthropology and other geological records. US National Science Foundation and The Smithsonian Institution's Human Origins Program, and DOSECC. Nov 17-20, 2005.

Brooks, A. S., Helgren, D. M., Cramer, J. S., Franklin, A., Hornyak, W., Keating, J. M., …Yellen, J. E. (1995). Dating and context of three Middle Stone Age sites with bone points in the Upper Semliki Valley, Zaire. *Science, 268*, 548-553.

Brown, F.H., & McDougall, I. (2012). Correlation of the KHS Tuff of the Kibish Formation to volcanic ash layers at other sites, and the age of early Homo sapiens (Omo I and Omo II). *Journal of Human Evolution, 63*(4), 577–585.

Clark, J. D., Beyene, Y., WoldeGabriel, G., Hart, W. K., Renne, P. R., Gilbert, H., …White, T. D. (2003). Stratigraphic, chronological and behavioural contexts of Pleistocene *Homo sapiens* from Middle Awash, Ethiopia. *Nature, 423*, 747-752.

Clark, J. D., & Brown, K. S. (2001). The Twin Rivers Kopje, Zambia, stratigraphy, fauna, and artefact assemblages from the 1954 and 1956 excavations. *Journal of Archaeological Science*, *28*(3), 305-330.

Clark, J. D., Oakley, K. P., Wells, L. H. & McClelland, J. A. C. (1947). New Studies on Rhodesian Man. *The Journal of the Royal Anthropological Institute of Great Britain and Ireland*, *77*(1), 7-32.

Claudot-Hawad, H. (2005). "Woman the Shelter" and "Man the Traveller": The Representation of Gender among the Tuaregs. Paper, Societies of Peace, 2nd World Congress on Matriarchal Studies, San Marcos and Austin, Texas USA, September 29-October 2, 2005, (accessed October 2014) http://www.second-congress-matriarchal-studies.com/claudot_hawad.html)

Clements, N. & Rialland, A. (2008). Africa as a phonological area. In B. Heine & D. Nurse (eds.), *A Linguistic Geography of Africa* (pp. 36-87). Cambridge: Cambridge University Press.

Cremaschi, M., Di Lernia, S., & Garcea E. A. A. (1998). Some insights on the Aterian in the Libyan Sahara: chronology, environment, and archaeology. *African Archaeological Review, 15*(4), 261-286.

Cruciani, F., Trombetta, B., Massaia, A., Destro-Bisol, G., Sellitto, D., & Scozzari, R. (2011). A revised root for the human Y chromosomal phylogenetic tree: the origin of patrilineal diversity in Africa. *The American Journal of Human Genetic,s 88*(6), 814-818.

Debénath, A. (1994). L'Atérien du nord de l'Afrique et du Sahara. *Sahara, 6*, 21-30.

de la Torre, I., Mora, R., Arroyo, A., & Benito-Calvo, A. (2014). Acheulean technological behaviour in the Middle Pleistocene landscape of Mieso (East-Central Ethiopia). *Journal of Human Evolution, 76*, 1-25.

d'Errico, F., Backwell, L., Villa, P., Degano, I., Lucejko, J. J., Bamford, M. K., … Beaumont, P.B. (2012). Early evidence of San material culture represented by organic artifacts from Border Cave, South Africa. *Proceedings of the National Academy of Sciences USA, 109*(33), 13214-13219.

d'Errico, F., Salomon, H., Vignaud, C., & Stringer, C. (2010). Pigments from the Middle Palaeolithic levels of Es-Skhul (Mount Carmel, Israel). *Journal of Archaeological Science, 37*(12), 3099-3110.

d'Errico, F., Vanhaeren, M., Barton, N., Bouzouggar, A., Mienis, H., Richter, D., ... Lozouet, P. (2009). Additional evidence on the use of personal ornaments in the Middle Paleolithic of North Africa. *Proceedings of the National Academy of Sciences USA*, *106*(38), 16051-16056.

Dimmendaal, G. J. (2001). Areal diffusion versus genetic inheritance: an African perspective. In A. Y. Aikhenvald & R. M. W. Dixon (eds.), *Areal diffusion and genetic inheritance: problems in comparative linguistics* (pp. 358-392). New York: Oxford University Press.

Dolgopolsky, A. (1998). *The Nostratic Macrofamily and Linguistic Paleontology*. Cambridge: McDonald Institute for Archaeological Research.

Dolgopolsky, A. (2008). *Nostratic Dictionary*. DSpace. Department of Archaeology, University of Cambridge. http://www.dspace.cam.ac.uk/handle/1810/196512.

Douka, K., Jacobs, Z., Lane, C., Grün, R., Farr, L., Hunt, C., ... Barker, G. (2014). The chronostratigraphy of the Haua Fteah cave (Cyrenaica, northeast Libya). *Journal of Human Evolution*, *66*, 39-63.

Douze, K. (2012). *Le Early Middle Stone Age d'Éthiopie et les changements techno-économiques à la période de l'émergence des premiers Homo sapiens* (Doctoral dissertation, Bordeaux 1).

Drake, N. A., Blench, R. M., Armitage, S. J., Bristow, C. S., & White, K. H. (2011). Ancient watercourses and biogeography of the Sahara explain the peopling of the desert. *Proceedings of the National Academy of Sciences USA, 108*(2), 458-462.

Eddy, F. W., Wendorf, F. & Associates. (1999). *An archaeological investigation of the central Sinai, Egypt*. Cairo: American Research Center in Egypt.

Ehret, C. (2011). Reconstructing ancient kinship in Africa. In N. Allen, H. Callan, R. Dunbar & W. James (eds.) *Early Human Kinship: From Sex to Social Reproduction* (pp. 200-231). Chichester UK: Wiley-Blackwell.

Feathers, J. K., & Migliorini, E. (2001). Luminescence dating at Katanda—a reassessment. *Quaternary Science Reviews*, *20*(5), 961-966.

Fernandes, V., Alshamali, F., Alves, M., Costa, M. D., Pereira, J. B., Silva, N. M., … Pereira, L. (2012). The Arabian cradle: mitochondrial relicts of the first steps along the southern route out of Africa. *American Journal of Human Genetics, 90*(2), 347-55.

Fernández Domínguez, E. (2005). *Polimorfismos de DNA mitocondrial en poblaciones antiguas de la cuenca mediterránea*. Tesis Doctorals. Universitat de Barcelona (online).

Fleming, H. C. (1987). Proto-Gongan Consonant Phonemes: Stage One. *Leo Reinisch. Werk und Erbe* (pp. 141-159). Wien: Verlag der Österreichischen Akademie der Wissenschaften,

Fleming, H. C. (1991). A New Taxonomic Hypothesis: Borean or Boralean. *Mother Tongue, 14,* n.p.

Fleming, H. C. (2002). Afrasian and Its Closest Relatives: the Borean Hypothesis. Paper, *Global Perspectives on Human Language: Scientific Studies in honor of Professor Joseph Greenberg*, April 25-27, 2002, Stanford University.

Fleming, H. C., Zegura, S. L., Harrod, J. B., Bengtson, J. D., & Keita, S.O.Y. (2013). The Early Dispersions of Homo sapiens sapiens and proto-Human from Africa. *Mother Tongue: Journal of the Association for the Study of Language in Prehistory XVIII*, 143-188.

Flood, J. (1990). *Archaeology of the dreamtime: The story of prehistoric Australia and its people*. New Haven: Yale University Press.

Fu, Q., Li, H., Moorjani, P., Jay, F., Slepchenko, S. M., Bondarev, A. A., ... Pääbo, S. (2014). Genome sequence of a 45,000-year-old modern human from western Siberia. *Nature, 514*(7523), 445-449.

Fu, Q., Mittnik, A., Johnson, P. L., Bos, K., Lari, M., Bollongino, R., … Krause, J. (2013). A revised timescale for human evolution based on ancient mitochondrial genomes. *Current Biology*, *23*(7), 553-559.

Gell-Mann, M., Peiros, I., & Starostin, G. (2009). Distant Language Relationship: The Current Perspective. *Journal of Language Relationship, 1*, 13-30.

Goldberg, P. (1986). Late Quaternary Environmental History of the Southern Levant. *Geoarchaeology, 1*(3), 225-244.

Gómez-Carballa, A., Olivieri, A., Behar, D. M., Achilli, A., Torroni, A., & Salas, A. (2012). Genetic continuity in the Franco-Cantabrian region: new clues from autochthonous mitogenomes. *PloS ONE, 7*(3), e32851.

Gonder, M. K, Mortensen, H. M, Reed, F. A, de Sousa, A, & Tishkoff, S. A. (2007). Whole-mtDNA genome sequence analysis of ancient African lineages. *Molecular Biology and Evolution 24*(3),757-68.

González-Ruibal, A. (2006). Order in a Disordered World: The Bertha House. *Anthropos,* 101, 379-402.

González-Ruibal, A & Fernández Martínez, V.M. (2007). Exhibiting Cultures of Contact, A museum for Benishangul-Gumuz. *Stanford Journal of Archaeology, 5,* 61–90.

Greenberg, J. H. (1955). *Studies in African Linguistic Classification*. New Haven: Compass Publishing Company. (Reprint of eight articles published in the *Southwestern Journal of Anthropology* from 1949 to 1954, with minor corrections.)

Greenberg, J. H. (1959). Africa as a linguistic area. In William R. Bascom and Melville J. Herskovitz (eds.), *Continuity and change in African cultures* (pp. 15-27). Chicago: Chicago University Press.

Greenberg, J. H. (1963). *The languages of Africa* (= Research Center in Anthropology, Folklore, and Linguistics Publications 25). Bloomington: Indiana University.

Greenberg, J. H. (1960). The general classification of Central and South American languages. In Anthony F. C. Wallace (ed.), *Selected Papers of the Fifth International Congress of Anthropological and Ethnological Sciences, 1956* (pp. 791-94). Philadelphia: University of Pennsylvania Press. (Reprinted in Greenberg 2005, 59-64.)

Greenberg, J. H. (1983). Some areal characteristics of African languages. In I. R. Dihoff (ed.), *Current approaches to African linguistics 1* (= Publications in African Languages and Linguistics 1). Dordrecht/ Cinnaminson: Foris, 3-22.

Greenberg, J. H. (1987). *Language in the Americas*. Stanford: Stanford University Press.

Greenberg, J. H. (2000). *Indo-European and Its Closest Relatives*. Stanford: Stanford University Press.

Greenberg, J. H. (2002). *Indo-European and Its Closest Relatives: The Eurasiatic Language Family, Volume 2: Lexicon*. Stanford: Stanford University Press.

Greenberg, J. H., & Ruhlen, M. (2007). *An Amerind Etymological Dictionary, Version 12*. Department of Anthropological Sciences, Stanford University.

Gregersen, E. A. (1972). Kongo-Saharan. *Journal of African Languages, 11*(1), 69-89.

Gruet, M. (1954). Le gisement mousterien d'El Guettar. *Karthago, 5*, 1-79.

Grün, R., Stringer, C. (2000). Tabun revisited: revised ESR chronology and new ESR and U-series analyses of dental material from Tabun C1. *Journal of human Evolution*, *39*(6), 601-612.

Grün, R., Stringer, C., McDermott, F., Nathan, R., Porat, N., Robertson, S., ... McCulloch, M. (2005). U-series and ESR analyses of bones and teeth relating to the human burials from Skhul. *Journal of Human Evolution*, *49*(3), 316-334.

Güldemann, T. (2008). The "Macro-Sudan belt": Towards identifying a linguistic area in northern Sub-Saharan Africa. In B. Heine & D. Nurse (eds.) *A Linguistic Geography of Africa* (pp. 151-185). , Cambridge: Cambridge University Press.

Güldemann, T. (2011). Proto-Bantu and Proto-Niger-Congo: macro-areal typology and linguistic reconstruction. In Osamu Hieda, Christa König and Hirosi Nakagawa (eds.) *Geographical typology and linguistic areas, with special reference to Africa* (pp. 109-141). Philadelphia: John Benjamins.

Harrod, J. B. (2010). Four Memes in the Two Million Year Evolution of Symbol, Metaphor and Myth. Presentation, *Exploratory Seminar on Comparative Mythology: Deep Reconstruction, October 6-8, 2010*. Radcliffe Institute for Advanced Study, Harvard University; co-sponsored with 4th Annual Conference, International Association for Comparative Mythology, Harvard University.

Harrod J. B. (2013). Already Out-of-Africa Before Out-of-Africa? Annual Reflection on Archaeology and Genetics. *Mother Tongue: Journal of the Association for the Study of Language in Prehistory, XVIII*, np.

Harrod J. (2014). Palaeoart at Two Million Years Ago? A Review of the Evidence. *Arts, 3*, 135-155.

Haslam, M., Roberts, R. G., Shipton, C., Pal, J. N., Fenwick, J. L., Ditchfield, P., Boivin, N., Dubey, A. K., Gupta, M. C., & Petraglia, M. (2011). Late Acheulean hominins at the Marine Isotope Stage 6/5e transition in north-central India. *Quaternary Research 75*(3): 670–682.

Hovers, E., Ilani, S., Bar-Yosef, O., & Vandermeersch, B. (2003). An early case of color symbolism: Ochre use by modern humans in Qafzeh Cave. *Current Anthropology, 44*(4), 491-522.

Hovers, E., Vandermeersch, B., & Bar-Yosef, O. (1997). A Middle Palaeolithic engraved artefact from Qafzeh Cave, Israel. *Rock Art Research, 14*(2), 79-87.

Hublin, J., & Klein, R. (2011). Northern Africa could also have housed the source population for living humans. *Proceedings of the National Academy of Sciences USA, 108*(28), E277

Hublin, J-J., & McPherron, S.P. (2012). *Modern Origins: A North African Perspective (Vertebrate Paleobiology and Paleoanthropology)*. Springer.

Hudjashov, G., Kivisild, T., Underhill, P.A., Endicott, P., Sanchez, J. J. Lin, A. A. … Forster, P. (2007). Revealing the prehistoric settlement of Australia by Y chromosome and mtDNA analysis. *Proceedings of the National Academy of Sciences of the USA, 104*(21), 8726-9730.

Hunley, K. L, Cabana, G. S, Merriwether, D. A, & Long, J. C. (2007). A formal test of linguistic and genetic coevolution in native Central and South America. *American Journal of Physical Anthropology, 132*(4), 622-31.

Illich-Svitych, V. A. (1971). *Tentative Comparison of the Nostratic Languages. Volume I: Introduction, Comparative Dictionary (b-K)*. Moscow: Nauka Publishers. Accessed http://starling.rinet.ru/babel.php?lan=en

Isabirye, D. (2010). *The mitochondrial DNA heritage of the Baganda, Lugbara and Acholi from Uganda*. PhD Dissertation, North-west University (Potchefstroom Campus), South Africa.

Jacobs, Z., Meyer, M. C., Roberts, R. G., Aldeias, V., Dibble, H., & El Hajraoui, M. A. (2011). Single-grain OSL dating at La Grotte des Contrebandiers ('Smugglers' Cave'), Morocco: improved age constraints for the Middle Paleolithic levels. *Journal of Archaeological Science* 38, 36341-3643.

Jacobs, Z., Roberts, R. G., Nespoulet, R., El Hajraoui, M. A., & Debénath, A. (2012). Single-grain OSL chronologies for Middle Palaeolithic deposits at El Mnasra and El Harhoura 2, Morocco: Implications for Late Pleistocene human–environment interactions along the Atlantic coast of northwest Africa. *Journal of Human Evolution, 62*(3), 377–394.

Jin, C-Z., Pan, W-S., Zhang, Y-Q., Cai, Y. J., Xu, Q. Q., Tang, Z. L., Wang, W., Wang, Y., Liu, J. Y., Qin, D. G. Edwards, R. L. & Cheng, H. (2009). The Homo sapiens Cave hominin site of Mulan Mountain, Jiangzhou District, Chongzuo, Guangxi with emphasis on its age. *Chinese Science Bulletin 54*(21): 3848-3856.

Jinam, T. A., Hong, L. C., Phipps, M. E., Stoneking, M., Ameen, M., Edo, J., & Saitou, N. (2012). Evolutionary history of continental Southeast Asians: "Early train" hypothesis based on genetic analysis of mitochondrial and autosomal DNA data. *Molecular Biology and Evolution*, *29*(11), 3513-3527.

Jones, S. C. & Pal, J. N. (2009). The Palaeolithic of the Middle Son valley, north-central India: Changes in hominin lithic technology and behaviour during the Upper Pleistocene. *Journal of Anthropological Archaeology 28*: 323-341.

Karmin, M. (2005). *Human mitochondrial DNA haplogroup R in India: dissecting the phylogenetic tree of South Asian-specific lineages* (Doctoral dissertation, MSc. thesis, University of Tartu, Estonia).

Kivisild, T., Reidla, M., Metspalu, E., Rosa, A., Brehm, A., Pennarun, E., … Villems, R. (2004). Ethiopian mitochondrial DNA heritage: tracking gene flow across and around the gate of tears. *American Journal of Human Genetics, 75*(5), 752-70.

Kobusiewicz, M. (1999). Excavations at Sinai-20, The Split Rock Site, Zarnoq Locality. In F.W. Eddy, F. Wendorf and Associates, *An archaeological investigation of the central Sinai, Egypt* (pp. 193-207). Cairo: American Research Center in Egypt.

Kobusiewicz, M., Schild, R., Bluszcz, A., & Wendorf, F., (2001). Reassessing Chronostratygraphic Position of the Split Rock Site, Sinai. In B. Gehlen, M. Heinen & A. Tillman (eds.), *Archäologishe Berichte 14, Band 1, Zeit Räume. Deutsche Gesellschaft für Ur- und Frühgeschichte*, 227-236.

Kumar, S., Padmanabham, P. B. S. V., Ravuri, R. R., Uttaravalli, K., Koneru, P., Mukherjee, P. A., … Rao, V. R. (2008). The earliest settlers' antiquity and evolutionary history of Indian populations: evidence from M2 mtDNA lineage. *BMC Evolutionary Biology*, *8*(1), 230.

Kumar, S., Ravuri, R. R., Koneru, P., Urade, B. P., Sarkar, B. N., Chandrasekar, A., & Rao, V. R. (2009). Reconstructing Indian-Australian phylogenetic link. *BMC Evolutionary Biology*, *9*(1), 173.

Kurashina, H. (1978). *An examination of prehistoric lithic technology in eastcentral Ethiopia*. PhD Dissertation, University of California, Berkeley

Liu, W., Jin, C-Z., Zhang, Y-Q., Cai, Y-J., Xing, S., Wu, X-J., Cheng, H., Edwards, R. L., Pan, W-S., Qin, D-G., An, Z-S., Trinkaus, E., & Wu, X-Z. (2010). Human remains from Zhirendong, South China, and modern human emergence in East Asia. *Proceedings of the National Academy of Sciences USA 107*(45): 19201-19206.

Liu, W., Wu, X., Pei, S., Wu, X. & Norton, C. J. (2010). Huanglong Cave: A Late Pleistocene human fossil site in Hubei Province, China. *Quaternary International 211*: 29-41.

Maji, S., Krithika, S., & Vasulu, T.S. (2008). Distribution of mitochondrial DNA macrohaplogroup N in India with special reference to haplogroup R and its sub-haplogroup U. *International Journal of Human Genetics*, *8*(1-2), 85-96.

Malyarchuk, B. A. (2004). Differentiation of the mitochondrial subhaplogroup U4 in the populations of Eastern Europe, Ural, and Western Siberia: Implication to the genetic history of the Uralic populations. *Russian Journal of Genetics*, *40*(11), 1281-1287.

Marks, A. E. (2009). The Paleolithic of Arabia in an Inter-regional Context. In M. D. Petraglia & J. I. Rose (eds.), *Evolution of Human Populations in Arabia: Paleoenvironments, Prehistory and Genetics* (pp. 295-308). Dordrecht: Springer.

Mayer, B-Y. D., Vandermeersch, B., & Bar-Yosef, O. (2009). Shells and ochre in Middle Paleolithic Qafzeh Cave, Israel: indications for modern behavior. *Journal of Human Evolution, 56*, 307–314.

McBrearty, S., & Brooks, A. S. (2000). The revolution that wasn't: a new interpretation of the origin of modern human behavior. *Journal of Human Evolution, 39*, 453-563.

McCarthy, R. C., & Lucas, L. (2014). A morphometric re-assessment of BOU-VP-16/1 from Herto, Ethiopia. *Journal of Human Evolution, 74*, 114-117.

McDougall, I., Brown, F. H., & Fleagle, J. G. (2005). Stratigraphic placement and age of modern humans from Kibish, Ethiopia. *Nature, 433*, 733-736.

Mehlman, M. J. (1987). Provenience, age and associations of archaic *Homo sapiens* crania from Lake Eyasi, Tanzania. *Journal of Archaeological Science, 14*(2), 133-162.

Mellars, P., Gori, K. C., Carr, M., Soares, P. A., & Richards, M. B. (2013). Genetic and archaeological perspectives on the initial modern human colonization of southern Asia. *Proceedings of the National Academy of Sciences USA, 110*(26), 10699-10704.

Mercader, J. (2002). Forest people: The role of African rainforests in human evolution and dispersal. *Evolutionary Anthropology: Issues, News, and Reviews*, *11*(3), 117-124.

Mercier, N., & Valladas, H. (1994). Thermoluminescence dates for the Paleolithic Levant. In O. Bar-Yosef & R. S. Kra (eds.), *Late Quaternary Chronology and Paleoclimates of the Eastern Mediterranean* (pp. 13-20). Tucson, Arizona: Radiocarbon and the Peabody Museum.

Mercier, N., Valladas, H., Forget, L., Joron, J. L., Vermeersch, P. J. M., Van Peer, P., & Moeyersons, J. (1999). Thermoluminescence dating of a Middle Palaeolithic occupation at Sodmein Cave, Red Sea Mountains (Egypt). *Journal of Archaeological Science, 26*(11),1339-1345.

Mercier, N., Valladas, H., Valladas, G., Reyss, J. L., Jelinek, A., Meignen, L., & Joron, J. L. (1995). TL dates of burnt flints from Jelinek's excavations at Tabun and their implications. *Journal of Archaeological Science*, *22*(4), 495-509.

Metspalu, M., Kivisild, T., Metspalu, E., Parik, J., Hudjashov, G., Kaldma, K., ... Villems, R. (2004). Most of the extant mtDNA boundaries in South and Southwest Asia were likely shaped during the initial settlement of Eurasia by anatomically modern humans. *BMC Genetics, 5*(26), e1-25.

Middleton, J. (1960). *Lugbara religion: ritual and authority among an East African people*. Oxford: Oxford University Press.

Middleton, J. (1979). Rites of sacrifice among the Lugbara, *Systèmes de pensée en Afrique noire* 4 (online accessed 2013).

Mijares, A.S., Détroit, F., Piper, P., Grün, R., Bellwood, P., Aubert, M., Champion, G, Cuevas, N., De Leon, A. & Dizon, E. 2010. New evidence for a 67,000-year-old human presence at Callao Cave, Luzon, Philippines. *Journal of Human Evolution 59*(1): 123-132.

Millard, A. R. (2008). A critique of the chronometric evidence for hominid fossils: I. Africa and the Near East 500–50ka. *Journal of Human Evolution*, *54*(6), 848-874.

Miller, R. A. (1971). *Japanese and the Other Altaic Languages*. Chicago: University of Chicago Press.

Miller, R. A. (1991). Genetic connections among the Altaic languages. In Sydney M. Lamb & E. Douglas Mitchell (eds.), *Sprung from Some Common Source: Investigations into the Prehistory of Languages* (pp. 293–327). Stanford: Stanford University Press.

Mishra, S, Chauhan, N, & Singhvi, A. K. (2013). Continuity of Microblade Technology in the Indian Subcontinent Since 45 ka: Implications for the Dispersal of Modern Humans. *PLoS ONE, 8*(7), e69280.

Negash, A., Brown, F., & Nash, B. (2011). Varieties and Sources of Artefactual Obsidian in the Middle Stone Age of the Middle Awash, Ethiopia. *Archaeometry, 53*(4), 661-673.

Nikolayev, S. L. (1991). Sino-Caucasian Languages in America. In V.V. Shevoroshkin (ed.), *Dene–Sino-Caucasian Languages* (pp. 42–66). Bochum: Brockmeyer.

O'Connell, J. F. and Allen, J. (2004). Dating the colonization of Sahul (Pleistocene Australia—New Guinea): a review of recent research. *Journal of Archaeological Science 31*(6): 835-853.

Pagani, L., Kivisild, T., Tarekegn, A., Ekong, R., Plaster, C., Gallego Romero, I., Ayub, Q., ... Tyler-Smith, C. (2012). Ethiopian genetic diversity reveals linguistic stratification and complex influences on the Ethiopian gene pool. *The American Journal of Human Genetics*, *91*(1), 83-96.

Pagel, M., Atkinson, Q. D., & Meade, A. (2007). Frequency of word-use predicts rates of lexical evolution throughout Indo-European history. *Nature, 449*(7163), 717-720.

Pala, M., Olivieri, A., Achilli, A., Accetturo, M., Metspalu, E., Reidla, M., ... Richards, M. B. (2012). Mitochondrial DNA signals of late glacial recolonization of Europe from near eastern refugia. *The American Journal of Human Genetics*, *90*(5), 915-924.

Palanichamy, M. G., Sun, C., Agrawal, S., Bandelt, H-J., Kong, Q-P., Khan, F., ... Zhang, Y-P. (2004). Phylogeny of mitochondrial DNA macrohaplogroup N in India, based on complete sequencing: implications for the peopling of South Asia. *The American Journal of Human Genetics*, *75*(6), 966-978.

Perego, U. A., Achilli, A., Angerhofer, N., Accetturo, M, Pala, M., Olivieri, A., …Torroni A. (2009). Distinctive Paleo-Indian migration routes from Beringia marked by two rare mtDNA haplogroups. *Current Biology, 19*,1, 1-8.

Pedersen, H. (1931). *The Discovery of Language: Linguistic Science in the Nineteenth Century.* Bloomington: Indiana University Press.

Perreault, C., & Mathew, S. (2012). Dating the Origin of Language Using Phonemic Diversity. *PLoS ONE, 7*(4), e35289.

Petraglia, M. D., Alsharekh, A., Breeze, P., Clarkson, C., Crassard, R., Drake, N. A., …Veall, M-A. (2012). Hominin Dispersal into the Nefud Desert and Middle Palaeolithic Settlement along the Jubbah Palaeolake, Northern Arabia. *PLoS ONE, 7*(11), e49840.

Petraglia, M. D., Alsharekh, A. M., Crassard, R., Drake, N. A., Groucutt, H., Parker, A. G., & Roberts, R. G. (2011). Middle Paleolithic occupation on a Marine Isotope Stage 5 lakeshore in the Nefud Desert, Saudi Arabia. *Quaternary Science Reviews, 30*(13–14), 1555–1559.

Petraglia, M. D., Haslam, M., Fuller, D. Q., Boivin, N., & Clarkson, C. (2010). Out of Africa: new hypotheses and evidence for the dispersal of Homo sapiens along the Indian Ocean rim. *Annals of Human Biology, 37*(3), 288-311.

Plaster, C. A. (2011). *Variation in Y chromosome, mitochondrial DNA and labels of identity in Ethiopia.* PhD thesis, The Centre for Genetic Anthropology, Department of Genetics Evolution, University College London.

Quam, R. M., & Smith, F. H. (1998). A reassessment of the Tabun C2 mandible. In T. Akazawa, K. Aoki and O. Bar-Yosef (eds.), *Neandertals and modern humans in Western Asia* (pp. 405-421). New York: Plenum Press.

Quintana-Murci, L., Quach, H., Harmant, C., Luca, F., Massonnet, B., Patin, E., … Behar, D. M. (2008). Maternal traces of deep common ancestry and asymmetric gene flow between Pygmy hunter-gatherers and Bantu-speaking farmers. *Proceedings of the National Academy of Sciences of the USA,* 105(5), 1596-601.

Ramos, J., Bernal, D., Domínguez-Bella, S., Calado, D., Ruiz, B., Gil, M. J., … Chamorro, S. (2008). The Benzú rockshelter: a Middle Palaeolithic site on the North African coast. *Quaternary Science Reviews, 27*(23), 2210-2218.

Ranaweera, L., Kaewsutthi, S., Tun, A. W., Boonyarit, H., Poolsuwan, S., & Lertrit, P. (2014). Mitochondrial DNA history of Sri Lankan ethnic people: their relations within the island and with the Indian subcontinental populations. *Journal of Human Genetics, 59*(1), 28-36.

Reyes-Centeno, H., Ghirotto, S., Détroit, F., Grimaud-Hervé, D., Barbujani, G., & Harvati, K. (2014). Genomic and cranial phenotype data support multiple modern human dispersals from Africa and a southern route into Asia. *Proceedings of the National Academy of Sciences, 111*(20), 7248-7253.

Reyss, J. L., Valladas, H., Mercier, N., Froget, L., & Joron, J. L. (2007). Application des méthodes de la thermoluminescence et des déséquilibres dans la famille de l'uranium au gisement archéologique d'El Akarit. *El Akarit. Un site archéologique du paléolithique moyen dans le sud de la Tunisie* (pp. 357-363). Paris : CulturesFrance - Éditions recherche sur les civilisations.

Richter, D., Moser, J., Nami, M., Eiwanger, J., & Midada, A. (2010). New chronometric data from Ifri n'Ammar (Morocco) and the chronostratigraphy of the Middle Palaeolithic in the Western Maghreb. *Journal of Human Evolution, 59*, 672-679.

Rink, W. J., Schwarcz, H. P., Ronen, A., & Tsatskin, A. (2004). Confirmation of a near 400 ka age for the Yabrudian industry at Tabun Cave, Israel. *Journal of Archaeological Science, 31*(1), 15-20.

Rito, T., Richards, M. B., Fernandes, V., Alshamali, F., Cerny, V., Pereira, L., & Soares, P. (2013). The first modern human dispersals across Africa. PloS ONE, *8*(11), e80031.

Roberts, R.G., Jones, R., Smith, M.A. (1990). Thermoluminescence dating of a 50,000-year-old human occupation site in northern Australia. *Nature 345*:153-156.

Rosa, A. & Brehm, A. (2011). African human mtDNA phylogeography at-a-glance. *Journal of Anthropological Sciences, 89*, 25-58.

Rose, J. I. (2010). New Light on Human Prehistory in the Arabo-Persian Gulf Oasis. *Current Anthropology, 51*(6), 849-883.

Rose, J. I. (2007). The Arabian Corridor Migration Model: archaeological evidence for hominin dispersals into Oman during the Middle and Upper Pleistocene. In *Proceedings of the Seminar for Arabian Studies* (pp. 219-237). Oxford: Archaeopress.

Rose, J. I., Usik, V.I., Marks, A.E., Hilbert, Y.H., Galletti, C., Parton, A., … Roberts, R. G. (2011). The Nubian Complex of Dhofar, Oman: An African Middle Stone Age Industry in Southern Arabia. *PLoS ONE, 6*(11), e28239.

Roset, J. P., & Harbi Riahi, M. (2007). *El Akarit: Un site archéologique du Paléolithique Moyen dans le Sud de la Tunisie*. Paris : CulturesFrance - Recherche sur les Civilisations.

Ruhlen, M. (1994a). *The Origin of Language: Tracing the Evolution of the Mother Tongue.* New York: John Wiley and Sons.

Ruhlen, M. (1994b). *On the Origin of Languages: Studies in Linguistic Taxonomy.* Stanford: Stanford University Press.

Sahle, Y., Hutchings, W. K., Braun, D. R., Sealy, J. C., Leah, E., Morgan, L. E, Negash, A., & Atnafu, B. (2013). Earliest Stone-Tipped Projectiles from the Ethiopian Rift Date to >279,000 Years Ago. *PLoS ONE, 8*(11), e78092.

Sawada, M. (2001-2003). Rethinking Methods and Concepts of Anthropological Studies on African Pygmies' World View: The Creator-God and the Dead. *African study monographs,* Supplementary issue, *27*, 29-42.

Scerri, E. M. (2012). The Aterian and its place in the North African Middle Stone Age. *Quaternary International, 300*, 111-130..

Scerri, E. M., Groucutt, H. S., Jennings, R. P., & Petraglia, M. D. (2014). Unexpected technological heterogeneity in northern Arabia indicates complex Late Pleistocene demography at the gateway to Asia. *Journal of Human Evolution, 75*, 125-142.

Schild, R. (1998). Comment on 'Van Peer, The Nile Corridor and the Out-of-Africa model: an examination of the archaeological record, 1998'. *Current Anthropology, 39* supplement, S134-S135.

Schild, R. (1999). Geomorphology, Stratigraphy and Lithology of Prehistoric Sites. In F.W. Eddy, F. Wendorf and Associates, *An archaeological investigation of the central Sinai, Egypt* (pp. 311-207). Cairo: American Research Center in Egypt.

Schlebusch, C. M., Skoglund, P., Sjödin, P., Gatteapille, L. M., Hernandez, D., Jay, F. ... Jakobsson, M. (2012). Genomic Variation in Seven Khoe-San Groups Reveals Adaptation and Complex African History. *Science*, *338*, 374-379.

Schwarcz, H. P., Blackwell, B., Goldberg, P., & Marks, A. E. (1979). Uranium series dating of travertine from archaeological sites, Nahal Zin, Israel. *Nature, 277*, 558-560.

Shea, J. J. (2008). The Middle Stone Age archaeology of the Lower Omo Valley Kibish Formation: Excavations, lithic assemblages, and inferred patterns of early Homo sapiens behavior. *Journal of Human Evolution, 55*, 448–485.

Shea, J. J., Fleagle, J. G., Brown, F., Assefa, Z., Feibel, C., McDougall, I., Bender, L., & Jagich, A. (2004). Archaeology of the Kibish Formation, Lower Omo Valley, Ethiopia. *Abstract of the Paleoanthropology Society Annual Meeting 2004.* Philadelphia.

Shipton, C., Clarkson, C., Pal, J. N., Jones, S. C., Roberts, R. G., Harris, C., ... & Petraglia, M. D. (2013). Generativity, hierarchical action and recursion in the technology of the Acheulean to Middle Palaeolithic transition: A perspective from Patpara, the Son Valley, India. *Journal of Human Evolution, 65*(2), 93-108.

Smith, J. R., Giegengack, R., Schwarcz, H., McDonald, M. M., Kleindienst, M., Hawkins, A., & Churcher, C. S. (2004). A reconstruction of Quaternary pluvial environments and human occupations using stratigraphy and geochronology of fossil-spring tufas, Kharga Oasis, Egypt. *Geoarchaeology, 19*(5), 1-34.

Smith, J. R., Hawkins, A. L., Asmerom, Y., Polyak, V., & Giegengack, R. (2007). New age constraints on the Middle Stone Age occupations of Kharga Oasis, Western Desert, Egypt. *Journal of Human Evolution, 52*, 690-701.

Smith, T. M., Tafforeau, P., Reid, D. J., Grün, R., Eggins, S., Boutakiout, M. & Hublin, J-J. (2007). Earliest evidence of modern human life history in North African early *Homo sapiens*. *Proceedings of the National Academy of Sciences USA, 104*(15), 6128-6133.

Soares, P., Alshamali, F., Pereira, J. B., Fernandes, V., Silva, N. M., Afonso, C., … Pereira L. (2012). The Expansion of mtDNA Haplogroup L3 within and out of Africa. *Molecular Biology and Evolution, 29*(3), 915-27.

Soares, P, Ermini, L, Thomson, N, Mormina, M, Rito, T, Röhl, A, Salas, A, Oppenheimer, S, Macaulay, V, & Richards, M. B. (2009). Correcting for purifying selection: an improved human mitochondrial molecular clock. *American Journal of Human Genetics*, *84*(6), 740-59.

Starostin, S. A. (1999). Subgrouping of Nostratic: Comments on A. Dolgopolsky's "Nostratic Macrofamily and Linguistic Paleontology". In *Nostratic: Examining a Linguistic Macrofamily* (pp. 137-156). Institute of Archaeological Research, Cambridge. Accessed http://starling.rinet.ru/babel.php?lan=en

Starostin, S. A. (2004-05a). Sino-Caucasian [comparative phonology]. Unpublished manuscript. The Tower of Babel: Evolution of Human Language Project. Accessed http://starling.rinet.ru/babel.php?lan=en

Starostin, S. A. (2004-05b). Sino-Caucasian [comparative glossary]. Unpublished manuscript. The Tower of Babel: Evolution of Human Language Project. Accessed http://starling.rinet.ru/babel.php?lan=en

Starostin, S. (2006). Borean tree diagram. http://starling.rinet.ru/images/globet.png.

Sukernik, R. I., Volodko, N. V., Mazunin, I. O., Eltsov, N. P., Dryomov, S. V., & Starikovskaya, E. B. (2012). Mitochondrial genome diversity in the Tubalar, Even, and Ulchi: contribution to prehistory of native Siberians and their affinities to Native Americans. *American Journal of Physical Anthropology*, *148*(1), 123-138.

Szabo, B. J., Haynes, C. V., & Maxwell, T. A. (1995). Ages of Quaternary pluvial episodes determined by uranium-series and radiocarbon dating of lacustrine deposits of Eastern Sahara. *Palaeogeography, Palaeoclimatology, Palaeoecology*, *113*(2-4), 227-242.

Taylor, N. J. (2011). The origins of hunting and gathering in the Congo basin: a perspective on the Middle Stone Age Lupemban industry. *Before Farming*, *1*, 6, 1-20.

Thangaraj, K., Chaubey, G., Singh, V. K., Vanniarajan, A., Thanseem, I., Reddy, A. G., & Singh, L. (2006). In situ origin of deep rooting lineages of mitochondrial Macrohaplogroup 'M' in India. *BMC Genomics*, *7*(1), 151.

Tishkoff, S. A., Gonder, M. K., Henn, B. M., Mortensen, H., Knight, A., Gignoux, C., … Mountain, J. L. (2007). History of click-speaking populations of Africa inferred from mtDNA and Y chromosome genetic variation. *Molecular Biology and Evolution, 24*(10), 2180-95.

Tishkoff, S. A., Reed, F. A., Friedlaender, F. R., Ehret, C., Ranciaro, A., Froment, A., … Williams, S. M. (2009). The Genetic Structure and History of Africans and African Americans. *Science, 324*(5930), 1035–1044.

Trinkaus, E. (1983). *The Shanidar Neandertals*. New York: Academic Press.

Trinkaus, E. (1984). Western Asia. In F.H. Smith & F. Spencer (eds.), *The Origins of Modern Humans: A world Survey of the Fossil Evidence* (pp. 251-293). New York: Academic Press.

Trinkaus, E. (2009). The human tibia from Broken Hill, Kabwe, Zambia. *PaleoAnthropology 2009*, 145-165.

Trombetti, Alfredo. (1905). *L'unità d'origine del linguaggio*. Bologna: Luigi Beltrami.

Valladas, H., Mercier, N., Joron, J. L., & Reyss, J. L. (1998). GIF laboratory dates for Middle Paleolithic Levant. In T. Akazawa, K. Aoki, & O. Bar-Yosef (eds.), *Neandertals and modern humans in western Asia* (pp. 69-76). New York: Plenum Press.

van de Loo, J. (1991). 'L'universo in miniatura.' *Nigrizia, 109*(10), 68-69.

Vandermeersch, B. (1981). *Les Hommes Fossiles de Qafzeh (Israel)*, Paris: CNRS Editions.

Van Driem, G. (2008). To which language family does Chinese belong, or what's in a name. In A. Sanchez-Mazas, R. Blench, M. D. Ross, I. Peiros & M. Lin (eds.), *Past Human Migrations in East Asia: Matching Archaeology, Linguistics and Genetics* (pp. 219-253). Routledge, London.

Vanhaeren, M., d'Errico, F., Stringer, C., James, S. L., Todd, J. A., & Mienis, H. K. (2006). Middle Paleolithic Shell Beads in Israel and Algeria. *Science, 312*, 1785-1788.

van Holst Pellekaan, S. M., Ingman, M., Roberts-Thomson, J., & Harding, R. M. (2006). Mitochondrial genomics identifies major haplogroups in Aboriginal Australians. *American Journal of Physical Anthropology. 131*, 282–94

van Oven, M., & Kayser, M. (2009). Updated Comprehensive Phylogenetic Tree of Global Human Mitochondrial DNA Variation. *Human Mutation, Mutation in Brief, 1039*(29), E386-E394. PhyloTree.org – mtDNA tree Build 15 (20 Jul 2011).

Van Peer, P. (1998). The Nile Corridor and the Out-of-Africa model: an examination of the archaeological record. *Current Anthropology, 39*, supplement, S115-S140.

Van Peer, P., Fullagar, R., Stokes, S., Bailey, R. M,, Moeyersons, J,, Steenhoudt, F,, ... Geus, F. (2003). The Early to Middle Stone Age transition and the emergence of modern human behaviour at site 8-B-11, Sai Island, Sudan. *Journal of Human Evolution, 45*, 187-193.

Vermeersch, P. M., Van Peer, P., & Paulissen, E. (2010). *Chert Quarrying, Lithic Technology, and a Modern Human Burial at the Palaeolithic Site of Taramsa 1, Upper Egypt* (Egyptian Prehistory Monographs). Leuven, NL: Leuven University Press.

Vermeersch, P. M., Van Peer, P., & Rots, V. (2005). A Middle Palaeolithic site with blade technology at Al Tiwayrat, Qena, Upper Egypt. *Antiquity, 79*, 305,1-7.

Walter, R.C., Buffler, R.T., Bruggemann, J.H., Guillaume, M.M., Berhe, S.M., Negassi, B.,... Gagnon, M. (2000). Early human occupation of the Red Sea coast of Eritrea during the Last Interglacial. *Nature, 405*, 65-69.

Watts, I. (2009). Red ochre, body-painting, and language: interpreting the Blombos ochre. In Rudolf Botha and Chris Knight (eds.), *The cradle of language* (pp. 63-92). New York: Oxford University Press.

Wendorf, F., Schild, R., Close, A. E., Schwarcz, H. P., Miller, G. H., Grün, R., ... McKinney, C. (1994). A chronology for the Middle and Late Pleistocene wet episodes in the Eastern Sahara. In O. Bar-Yosef & R. S. Kra (eds.), *Late Quaternary chronology and paleoclimates of the Eastern Mediterranean* (pp. 147-168). Ann Arbor, MI: Braun-Brumfield.

Westermann, D. (1911). *Die Sudansprachen: Eine sprachvergleichende Studie* (= Abhandlungen des Hamburgischen Kolonialinstituts 3). Hamburg: L. Friederichsen.

Westermann, D. (1935). Charakter und Einteilung der Sudansprachen. *Africa, 8*(2), 129-148.

Westermann, D. (1949). *Sprachbeziehungen und Sprachverwandtschaft in Afrika* (= Sitzungsberichte der Deutschen Akademie der Wissenschaften zu Berlin, Philosophisch-historische Klasse, Jahrgang 1948,1).

White, T. D. (1986). Cut marks on the Bodo cranium: A case of prehistoric defleshing. *American Journal of Physical Anthropology, 69*, 503-509.

Witzel, E. J. M. (2011). *The Origins of the World's Mythologies*. New York: Oxford University Press.

Xing, J., Watkins, W. S., Hu, Y., Huff, C. D., Sabo, A., Muzny, D. M., ...Yu, F. (2010). Genetic diversity in Indian the inference of Eurasian population expansion. *Genome Biology* 11(11), R113.

Yellen, J. E., Brooks, A. S., Cornelissen, E., Mehiman, M. J., & Stewart, K. (1995). A middle stone age worked bone industry from Katanda, Upper Semliki Valley, Zaire. *Science, 268*, 553-556.

Yellen, J., Brooks, A., Helgren, D., Tappen, M., Ambrose, S., Bonnefille, R., ... Stewart, K. (2005). The archaeology of Aduma Middle Stone Age sites in the Awash Valley, Ethiopia. *PaleoAnthropology 10*, 25-100.

Yunusbayev, B., Metspalu, M., Järve, M., Kutuev, I., Rootsi, S., Metspalu, E., ... & Villems, R. (2012). The Caucasus as an asymmetric semipermeable barrier to ancient human migrations. *Molecular Biology and Evolution, 29*(1), 359-365.

The Kinship Term *KUKU ~ KOKO ~ KAKA* in the American Indigenous Languages, the Amerind Hypothesis, and the Dravidian Kinship System[1]

Part I: Linguistic Study

Alain Matthey de l'Etang & Pierre J. Bancel [f]

Abstract. The study of kinship terminologies in American indigenous languages unequivocally shows that the terms *KOKO ~ KUKU ~ KAKA* 'MB, EF' or 'GM, FZ' are present in most regions of the double continent, in a vast majority of linguistic families as well as in many languages not yet classified or considered isolated. The distribution of this term cannot ultimately be explained in terms of areal diffusion, as the continental distribution of the linguistic reconstructions demonstrates, but points to an ancestry within each linguistic family with time depths of several millennia. The antiquity of this term, its global distribution across the Americas, its phonetic properties and its semantic consistency, along with the fact that, in all probability, the initial peopling of the Americas was the result of a major colonizing event by a single source population, are consistent with the hypothesis that one or several *KOKO ~ KAKA* terms were present in the kinship lexicon of this founding population, which on genetic, archeological and geographical grounds may have entered in America as early as 16,000 years ago.

Abbreviations: P 'parent,' **G** 'grand,' **M** 'mother,' **F** 'father,' **Z** 'sister,' **B** 'brother,' **U** 'uncle,' **A** 'aunt,' **E** 'spouse,' **e/y** 'elder/younger,' **W** 'wife,' **H** 'husband,' **C** 'child,' **S** 'son,' **D** 'daughter,' **inL** '-in-law,' **Sib** 'sibling.' Other relationships are obtained by combination of these primary symbols: **MB** 'mother's brother,' **GM** 'grandmother,' etc. The symbols (♀) and (♂) found before kin type abbreviations indicate the sex of the person speaking; **ad.**: vocative or address term; **ref.**: referential term; **AM & PB**: Matthey de l'Etang & Bancel; **EHL**: Evolution of Human Languages project (ehl.santafe.edu/intro1.htm); **ASJP**: Automated Similarity Judgment Program (http://wwwstaff.eva.mpg.de/~wichmann/ASJPHomePage.htm); **R**: 'reconstruction done by'; **LDRC**: University of Alberta, Language Documentation Research Cluster; **P-**: 'Proto-'; **dial.**: dialect.

1. Presentation

Our first goal in this paper is to make a general appraisal of the distribution in the Americas of a term long and well known to Amerindianists, which can be labeled *KOKO* and has the general semantic scope of GF, MB, EF for the masculine side and GM, FZ and EM for the feminine side. Our second goal is to give an explanation for its amazing distribution, which was first remarked on

---

[1] This paper (including the anthropological developments) was presented at the annual meeting of the American Anthropological Association (kinship session) held in San Francisco on November 14-18, 2012.

[f] Association d'études linguistiques et anthropologiques préhistoriques, Paris. First author's mail: <a.matthey@free.fr>.

by von Martius (1867a & b) during the first half of the 19th century in South America[2], and by other authors at the turn of the 20th century for Meso-America. Moreover, more than a hundred years after von Martius, Landar (1977: 423) stated that, although uncharted, *"The areal distribution of* koko-*type words for kinmen ... will involve North as well as South America."* Still more recently, Ruhlen (1994b: 218-19) confirmed Landar's claim, arguing in favour of the Proto-Amerind antiquity of *KAKA*, and furthermore proposing a joint Amerind-Eurasiatic etymology: Amerind **(k)aka* eB, eZ = Eurasiatic **aka* eB. The same year, Ruhlen (1994a: 122-124) published the global etymology *kaka* 'U, eB,' showing that this root was not only widespread in the Americas, but also in Eurasia and Oceania, a distribution that we later widened to many more families from these areas as well as from Africa, Southern Asia, Australia, and New Guinea. As a result, *KAKA* or *KOKO* words are attested in most language families worldwide, with reflexes in more than 1,300 languages in our database of some 2,400 kinship terminologies (Bancel & Matthey de l'Etang 2002; Bancel, Matthey de l'Etang & Bengtson 2011; Matthey de l'Etang & Bancel 2002, 2008; Matthey de l'Etang, Bancel & Ruhlen 2011)[3].

More recently, Dixon & Aikhenvald (1999: 8), Aikhenvald (2002: 294), Dixon (2004a: 13, 2004b: 17), apparently ignoring Ruhlen's publications, explained the South American lowland (Amazonian) distribution of *KOKO* in terms of areal diffusion. The two authors believe that what triggers such a chain of borrowings is the need, for societies practising intermarriage, to agree on the term referring to the father-in-law (EF), hence *KOKO*. Even if it is difficult to figure out exactly what their *"across-Amazonia"* distribution means in terms of the languages concerned, their conclusions oblige us to reconsider the distribution, meaning and origin of this term. Do *KOKO* terms found all over the Americas, beyond the particular destinies of their multiple forms, ultimately descend from a common source, from multiple sources, or are they the result of some gigantic cascade-borrowing process? In order to reach convincing answers, we will then specify the distribution of *KOKO* terms in various language families of South, Central and North America.[4] This will confirm, amplify and specify the presentations given by Ruhlen in 1994.

Our third goal is to unveil the nature of the kinship system of the group or groups that first reached the American Continent. This particular issue will be the object of the forthcoming second (anthropological) part of our paper.

Sections 2-4 present the geolinguistic distribution of *KOKO* forms and illustrate how reconstructions have already been worked out by linguists in a substantial number of linguistic families all over the Americas, some of which are at a time depth of over 5,000 years.[5] We also present potential cognate sets from other language families. In doing so, we will bring to light some deviant forms within a number of cognate sets, explaining them by phonetic or morphological properties, or sometimes by their probable borrowing from languages belonging to other linguistic stocks. Section 5 presents a synthesis of the reconstructions. Section 6 discusses the origin of *KOKO* in the Americas linking linguistic and genetic studies. Our conclusion will be that the distribution of these terms with highly consistent phonetic forms and meanings in numerous reconstructed proto-languages of the whole continent definitely precludes that they might have resulted from cascade borrowings, especially as most reconstructed families display neatly individualized forms that have consistently been preserved in an overwhelming majority of their respective descendants, to the exception of a small number of easily identifiable borrowings. This distribution is consistent and best explained by *KOKO* being a retention from Proto-Amerind, a hypothetical ancestral language, nowadays substantially supported both linguistically and genetically.

---

[2] Von Martius, while pointing out the distribution of forms such as *ghuk* or *koko* across the continent – most of the time referring to the uncle or the mother's brother –, classified languages using this term in what he called the *"coco or guck group"* (Martius 1867a: 346-347).

[3] We have now collected data from about 3,400 languages and dialects. See our database at language-kinship.org.

[4] Languages have been arrayed according to the classifications adopted by Campbell (2012) for South America, by Kaufman in the *Atlas of the World's Languages* (2nd edition 2007) for Middle (Meso)-America, and by Golla, Campbell, Mithun, Mixco and Goddard, in the same atlas, for North America.

[5] The dates given for proto-languages are taken from Kaufman (1994 [2nd edition 2007]), from the Automated Similarity Judgment Program consortium (http://wwwstaff.eva.mpg.de/~wichmann/ASJPHomePage.htm) or from the authors referenced in the appropriate columns of our tables.

2. The kinship term *KOKO* in South America

The first records of *KOKO* kinship terms in South America date from the 16th and 17th centuries: Tarascan *cucu* [kuku] GM (Gilberti 1559), Quechua *caca* [kaka] MB (Santo Tomás 1560a & b), Island Carib *neùkêcayem* [nø-kəka-jɛm] 'FZH, MZH, MB' (Breton 1665, 1666; with a 1st person possessive prefix *nø-*), Kipea (Kariri) *i-cucú* [kuku] MB (Mamiani 1698: 197), Muisca (Chibcha) *caca* [kaka] GM (anonymous manuscripts dating back to the beginning of the 17th century: ms. 158, voc. fol. 9r, at the National Library of Colombia, Bogota, and ms. 2923, fol. 4r, at the library of the Palacio Real, Madrid). The first author to mention the existence of the term *KOKO* on a large scale was von Martius (1867a & b). Others authors since then (Koch-Grünberg 1911; de Créqui-Montfort & Rivet 1921, 1922; Rivet 1948; de Goeje 1909, 1946; Schuller 1928; Dixon 2004a & b) collected *KOKO* terms in various linguistic families such as Arawakan (Maipuran), Takanan, Panoan, Cariban, Chibchan, Misumalpan, Arawan, etc.

2.1. The geographic and linguistic distribution of KOKO *in South America*

Campbell's (2012) classification of South American indigenous languages comprises 108 language units: 53 families comprising at least 2 languages, and 55 language isolates. Arraying our data according to this classification does not mean that we necessarily adhere to its preamble, but its conservative formulation allows us to present data without embarking into unnecessary controversy.

KOKO forms are found in 53 units out of the 108 considered in Campbell (2012): 35 families out of 53 comprising at least 2 languages, and in 18 language isolates out of the 30 for which we have data (63.9% of groupings).[6] The families where one or several *KOKO* forms have previously been reconstructed or suggested are, in the order where they are presented below: Arawakan (Maipuran), Pano-Takanan (Panoan, Takanan), Arawan, Cariban, Guahiboan, Chibchan, Quechuan, Aymaran, Mondé, Nambikuáran, and Karirian. We also present the Lencan and Misumalpan material (not included in Campbell 2012) in section 2, because there is convincing evidence that both these language families are related to Chibchan (Constenla Umaña 2012).[7] The other language units displaying *KOKO* forms are presented in subsection 2.5.

---

[6] This wide distribution of *KOKO* may be underestimated, as this word is primarily an address form, a category sometimes neglected by field linguists and anthropologists.

KOKO forms have been looked for in kinship terminologies from some 407 languages and dialects of South America[8] and found in all the geolinguistic units defined by Kaufman (2007) (map1):[9] North-Western Western Amazonia, Northern Foothills, Andes, Southern Foothills, the Cone, Eastern Brazil, North-Eastern Brazil, and Central Amazonia. The Lenca-Misumalpan forms present in Nicaragua, Honduras, and El Salvador are shown on map 2.

2.2. The phonetic series KOKO *in South American linguistic families*

Our list of potential cognates has been established taking *KOKO ~ KUKU* as a starting point, because of the obvious and long-known existence of such forms in South American indigenous languages. We admitted to our list kinship terms presenting no or almost no phonetic difference with this *KOKO ~ KUKU* model, as well as forms displaying vocalic differences, like *kuka*, *koka* or *kaka*. We also admitted non-reduplicated or partially reduplicated forms like *ghuk*, *uka*, *gu*, *ko*, and finally forms likely to display a phonetic evolution like *axa*, *axu*, etc., with velar fricatives most likely reflecting an earlier velar stop *k* or *g*. Finally, we also took into account forms prefixed with a 1st person singular possessive, like Curripaco (Arawakan) *nu-kui-ru* 'my FZ', or *nu-ki-ri* 'my MB', because their roots can be easily identified and related to their appellative counterparts *kuuku* FZ and *kiiki* MB. Out of the 407 kinship terminologies under examination, 300 (73%) exhibit 448 likely cognates.[10] Table 1 recapitulates the various subtypes encountered. This phonetic series is remarkably consistent and homogeneous, as 188 out of 448 reflexes (42%) are CVCV (*kVkV*) forms. Among these, reduplicated *koko* and *kuku* forms are predominant.

Table 1. Statistical distribution of phonetic forms

Phonetic forms	*KOKO*	*KUKU*	*KAKA*	**Other *kVkV* forms**	**Other forms**	**Total**
Number	57	41	25	65	260	448
Percentage	12.7	9.2	5.6	14.5	58	100

---

[7] Constenla Umaña (2012: 418-419) dates the split of the 'Lenmichi' (Lenca-Misumalpan-Chibchan) phylum between Proto-Chibchan and Proto-Lenca-Misumalpan to 9,726 BP ±1,105y. The "Lenmichi" homeland is posited in Honduras (Adelaar 2012: 419).

[8] This figure includes the Lenca-Misumalpan languages. Out of these 407 languages, 335 dialects and languages are still alive (299 languages, 36 dialects).

[9] Kaufman's (2007) classification arrays languages and language families along geolinguistic regions.

[10] A language can exhibit more than one reflex of *KOKO* referring to one or several relationships. Whenever a language displays variants of *KOKO* to refer to the same relationship, only one has been taken into consideration for our general count. In some cases, one reflex can also refer to both feminine and masculine relationships: GP = GF and GM. Such reflexes are counted twice when dealing with the semantic series.

2.3. The reconstruction of KOKO *in South American linguistic families*[11]

Speaking about the presence of the form *koko* in every language of the Arawan family, Dixon (2004b: 17) says that *"a similar form is found in many other Amazonian languages [from other families, AM & PB], constituting a clear areal feature. It could have been borrowed from anyone of a number of sources – in Proto-Arawan or else into individual languages."* At the same time, the same author also says that *"the Proto-Arawá term is koko"* (Dixon 2004a: 13). The idea which emerges from his remarks is that the distribution of *KOKO* terms throughout Amazonia might result from an ongoing but already ancient process.

But can one maintain such an idea, if similar forms can be reconstructed at the Proto-language stage of other indigenous language families of America – some of them very ancient –, especially when their respective homelands are far from one another, some of them even far from Amazonia?

We will review and discuss the reconstructions already achieved by previous authors, and will also propose obvious cognate sets from other language families (section 2.4.).

2.3.1. Reconstruction of KOKO *in Proto-Arawakan (Maipuran): before 3,000 BP?*

The Arawakan languages. (Map: Wikipedia).

The Proto-Arawakan homeland is currently posited between the Upper Amazon (Solimões) in Brazil and the Middle Orinoco in Venezuela (Heckenberger 2002: 103; see Map 1). The initial split of Proto-Arawakan is generally estimated to date back to before 3,000 BP (Heckenberger 2002: 106-9, quoting Noble 1965[12]), while Kaufman (2007: 65) reports Swadesh's dating of 4,500 BP. At the time of European contact, this family occupied a territory spanning from the Bahamas to Northern Argentina, that is to say a space extending far beyond the Amazonian Basin. According to Heckenberger (2002: 107), its maximal extent was reached by ca. 500 AD. Payne (1991: 413, 424) reconstructed two *KOKO* forms in Proto-Arawakan: **ahku(-ro)* A, EM, and **kuhko* U, EF, based on regular sound correspondences.

- **Proto-Arawakan **kuhko* U (MB), EF** (Payne's cognate sets numbered from **1** to **13** in boxed table 2).

A number of remarks have to be made regarding Payne's reconstruction of **kuhko* U, EF which is founded on a contrast between *u* and *o*, only pervasive in 3 languages belonging to 3 different branches of Arawakan (1991: 476): Terena (Southern Arawakan), Chamicuro (Chamicuro), and Guajiro (Caribbean Arawakan).

---

[11] When two different time depths are mentioned between brackets, the first one is from glottochronology, the second one from ASJP. When only one date BP is mentioned, round numbers are from glottochronology, the others from ASJP.

[12] Heckenberger's acception of Arawakan matches Aikhenvald's (1999a) classification and includes Southern Arawak, Paressi-Xingu, South-Western Arawak, Campa, Amuesha, Chamicuro, Rio Branco, Palikur, Caribbean, and Northern Amazonian. This grouping corresponds to Noble's (1965) Arawakan subgroups less the Arawan and Uru-Chipaya branches. It also matches Payne's (1991) Arawakan (Maipuran). According to Noble, Arawakan (Maipuran) is just one of the seven branches derived from Proto-Arawakan. It does not include Taino, Chamicuro, Amuesha and Apolista. Noble lexicostatistically estimated the split of Proto-Arawakan into its seven branches ca. 3300 BP (1965: 111), and that of its Arawakan (Maipuran) branch ca. 2500 BP (1965: 109-111).

Timote Cuica *kuxioj* GM
P-Chibchan *_kaka_ F; *_gaka_ EM, GM, (# _kuku_ EF, MB) Sapé *kohai* GF
P-Guahiboan *_axu-_ EF, MB; *_akwe_ GM Warao *ku* MB
Sáliba *koko* man; Puinave hika Au; **P-Cariban #_koko_ GF, MB, EF, #_kuku_ GM, FZ, EM**

Paez *kahka* MB, MZH Hupda *ʔuw* GF, *ʔɔh* GM
P-Arawakan (Maipuran) *-_kuhko_ EF, MB, FZH
***-_ahku_ EM, FZ, MBW**

Taushiro *'ukkuk* Z
Urarina *ka-kaun* FZ

Mundurucú *kötköt* MZS

P-Arawan *_koko_ EF, MB
P-Panoan *_koka_ EF, MB **P-Mondéan #_ko(t)ko(t)_ MB, EF** **P-Karirían #_kuku_ MB**

P-Quechuan *_kaka_ GF, MB, EF Purubora *kɔkɔ* 'tio' Trumai *koko* MZ
Jaqaru *kaka* U **P-Takanan *_kuku_ MB**
Kapixaná *keke* GM; Sabanê *kooka* FF, FM, EF, MB Kamakan *gkoonk* GF, MB
Canichana *eu-axa* my GF **P-Nambikuáran *_kṵ̃ka-_ EF, MB**

Maxakalí
Chamacoco *–okok* EF *xuxjã* GF, EF, MB, *xukux* GM, EM, FZ
Western Mascoy *koko* MB
Macá *kewket* EF

P-Southern Jê *_kãke_ eB
Lule *kue* FZ ***_kakrã_ MB, EF**

Mapuche *kuku* FM

Tehuelche *qon* GM, *koka* F, dueño

Kaweskar *xojko-* old woman, *kjewo-* GM

Selknam *hoʔo* GF, *hohoʔnh* GM

Map 1. Approximate locations of _koko_ across South America. Proto-forms (P-) appear in bold.

In order to explain the lack of contrast in the other branches, the author states (1991: 478-79) that: *"It must have been the case that the contrast between the rounded vowels in Proto-Maipuran carried a low functional load and was already beginning to be eroded."*[13] Besides, the presence of consonant *-h-* in the form reconstructed by Payne is based on the existence of *h* in the Chamicuro word *kohka* MB, which is certainly an exaggerated importance given to a single language, while Payne does not document regular loss of word-internal **h* in any other Arawakan language. As a result, we propose that Chamicuro be considered irregular, a more economical solution than positing a Proto-Arawakan root **kuhko* and (implicitly) assuming that all languages but Chamicuro are irregular in having lost it[14]. Another remark concerns the Campa languages where the kin term's second vowel depends on the sex of the speaker (*-o-* is feminine in Arawakan, while *-i-* is masculine). This remark could perhaps also apply to the North Amazonian languages, whose MB terms of address regularly display a second vowel *i*. Other comments will be found in the appropriate column of table 2. We need to mention finally the rather unexpected replacement of the back vowel *o* or *u* by a high front vowel *-i-*, found in South Arawak (Mučoxeone *ni-kiko* U, Bauré *-kik* MB, EF), which is explained by Payne (1991: 477, 479) by the regular correspondence – at least in Bauré – **u > i*. All that being said, these remarks are unlikely to cast doubt on the validity of the Proto-Arawakan root **kuko* U, EF, if one admits the contrast between *u* and *o*. In any case, the bulk of evidence presented in the following augmented cognate sets[15] warrants beyond any reasonable doubt that, in accordance with what occurs in a great many regions of the Americas, there existed a root form for MB and EF in Proto-Arawakan, presenting a reduplicated phonetic shape **koko* ~ **kuku*, with the possibility of the final vowel *i* indicating a male speaking.

➢ Comparative data for MB terms in Arawakan including possessive forms[16]

Southern Arawakan 1. Terena *eúko* MB, EF; Kinikinau *ēukɔ* (♂)U; **2.** Ignaciano (Moxo) *nekuka* U; Mučoxeone *nikiko* U; **3.** Bauré *-kik* MB, EF;

Southwestern Arawakan Piro *nukoxiru* MB, FZH, HF; **4.** Piro *koko* ad. MB, FZH, EF; Apuriná *ukõkuru* U; Cuniba *kuku* MB, EF; Cushichineri *koko* U; Kanamaré *ghughu* MB; Mashko *kokoa* U; Sirineiri *kokoa* U;

Paresi-Xingu Yawalapití *kukujú* (♀)MB; Mehináku *kuku* (♀)MB; **5.** Paresí *koko-re* U, EF; Saraveca *koko-re-ixi* U;

---

[13] Payne (1991: 478) attributes the final /u/ in Palikur *kuku* to some old unreliable source, and infers that it "should have been /o/."

[14] Alternatively, the Chamicuro form *kohka* MB, EF can be explained as a loan from Panoan *koka* MB, EF.

[15] We did not retain a number of *KOKO* terms such as Baré *nukáka* MB, *nukāka* EF, and Mandawaka *kaka* MB, EF or even Island Carib *nøkəka-yɛm* in our cognate set, in the absence of any supporting evidence for a phonetic correspondence **u/o > a*. These terms could have been borrowed from other languages, although we have no direct evidence for this. Let us also mention that the loss of (initial) **k* in Terena (and Kinikinau) is only supported by five examples (Payne 1991: 441). Besides, together with the root form, some cognates also display the 1sg possessive prefix, reconstructed by Matteson (1972: 164) as **n(V)-*, and by Payne (1987: 62) as **nu-*.

[16] Languages are arrayed according to Aikhenvald's classification (1999a: 67-71). Together with the root form, cognates also display the 1sg possessive prefix (see note 15), as well as the masculine relative suffix reconstructed as **-ri* by Matteson (1972: 164) and Payne (1987: 63).

Table 2. Arawakan *kVkV* forms (MB)

Languages			***k***	***V***	***k***	***V***		**Denotata**	**Linguistic comments**
1 Terena		*e*		*u*	*k*	*o*		MB, EF	Payne 1991: **u* > *u*, **o* > *o*
2 Moxo	*n*	*e*	*k*	*u*	*k*	*a*		U	Payne 1991: **u* > *u*, **o* > *a*, or < Panoan *kuka*?
Mučoxeone	*n*	*e*	*k*	*i*	*k*	*o*		U	**u* > *i* ?
3 Baure		-	*k*	*i*	*k*			MB, EF	Payne 1991: **u* > *i*
4 Piro			*k*	*o*	*k*	*o*		MB, EF, FZH	Payne 1991: **u* > *o*
Cuniba			*k*	*u*	*k*	*u*		MB, EF	
Cushichineri			*k*	*u*	*k*	*u*		U	
Kanamaré			g^h	*u*	g^h	*u*		MB	
Mashko, Sirineiri			*k*	*o*	*k*	*o*	*a*	U	
Apuriná		*u*	*k*	*õ*	*k*	*u*	-	U	
Yawalapiti			*k*	*u*	*k*	*u*	-	MB	
Mehinaku			*k*	*u*	*k*	*u*		MB	
5 Paresi			*k*	*o*	*k*	*o*	-	U, EF	Payne 1991 **u* ~ **o* > *o*
Saraveca			*k*	*o*	*k*	*o*	-	U	
Ashaninka (Campa)			*k*	*o*	*k*	*o*	-	(♀)MB	Final -*o* feminine marker?
Ashaninka (Campa)			*k*	*o*	*k*	*i*		(♂)MB	Final -*i* masculine marker?
Ashaninka (Tambo dialect)			*k*	*ṓ*	*k*	*o*		(♀)MB	Final -*o* feminine marker?
Ashaninka (Tambo dialect)			*k*	*o*	*k*	*i*		(♂)MB	Final -*i* masculine marker?
6 Ashéninka			*k*	*ō*	*k*	*o*		(♀) MB ad.	Final -*o* feminine marker?
Ashéninka			*k*	*o*	*k*	*i*		(♂) MB ad.	Final -*i* masculine marker?
Matsiguenga			*k*	*o*	*k*	*i*		MB, EF ad.	Final -*i* masculine marker?
Nomatsiguenga			*k*	*o*	*k*	*í*		(♂)MB, EF ad.	Final -*i* masculine marker?
7 Chamicuro			*k*	o^h	*k*	*a*		MB, EF ad.	Payne 1991: **o* > *a*, or < Panoan *koka*?
Manao			g^h	*ō*	*k*	*o*		MB	
Wainuma			g^h	*o*	x^h	*o*	*i*	MB	
Waraicu			g^h	*u*	*k*			MB	
8 Palikur			*k*	*u*	*k*	*u*		MB, EF ad.	Payne 1991: **u* > *u*, **o* > *u*
Marawa				*u*	*k*	*í*		U	-*i* masculine marker?
9 Yucuna				*o*	*k*	*u*		MB, EF	Payne 1991: **u* > *o*,
10 Cabiyari		*ā*			*k*	*u*		MB, EF	**u* > *u*, or < Proto-Arawakan *-*aku* FZ?
Resigaro			k^h	*íi*	*g*	*í*		MB ad.	-*i*- masculine marker?
Achagua			*k*	*ū́*	*w*	*i*		U ad.	-*i*- masculine marker? See *kúú* FZ
11 Piapoco			*k*	*u*		*í*		MB, EF ad.	-*i* masculine marker? See *kùu* FZ
Curripaco			*k*	*ii*	*k*	*i*		MB, EF ad.	-*i*- masculine marker? See *kuuku* FZ
Baniwa	*n*	*ú*		*ú*	*k*	*u*		MB	
12 Tariana	*n*	*u*	k^h			*í*		MB, EF ad.	Payne 1991: **u* > *i*

Guarequena	*n*	*u*	*k*		*k*	*o*	MB	
Guinau	*n*	*ú*		*ú*	*k*	*u*	EF	
13 Yavitero	*n*	*ó*		*ó*	*k*	*o*	MB	**u* > *o*?

Campa[17] Axininca *kokoini* ad. (♀)MB, HF, (♀) FZH, *koNki* ad. (♂)MB, WF, (♂)FZH, *no-koNkiri* my WF, *no-koNkit^{h}oriri* my distant U; Asháninka *-koNkiri* MB, EF, *kokoxi* (♀)MB, my (♂)EF; Asháninka (Tambo dialect) *kṓko* my MB (ego ♀), HF *koki* my MB (ego ♂), WF; Ashéninka (Apurucayali) *nokoNkiri* my MB, my EF, my FZH; **6.** Ashéninka (Apurucayali) *koko-ini* (*-ini* = masc. suffix) (♀)U, HF, (♀)FZH (ad.), *koNki* ~ *koki*, ad. (♂)U, WF, (♂)FZH; Ashéninka (Ucayali) *nokoNkiri* my MB (ego ♂), my EF (ego ♂), my FZH (ego ♂); **6.** Ashéninka (Ucayali) *koko-ini* (♀)U, HF, (♀)FZH (ad.), *koNki* ~ *koki* ad. (♂)U, WF, (♂)FZH; Ashéninka Perené *koki* ad. MB (ego ♂), WF, *kooko* ad. MB (ego ♀), *nokonkiri* MB, EF; Caquinte *kōNk-ini* 'tio'; Nomatsiguenga *nokoNgiri* my MB, my EF, my FZH, *koki* (ad.) (♂)MB, (♂)EF; Matsiguenga *nogókine* my MB, EF, *koki* ad. MB, EF; Nanti *igoNkirite* his WF;

Amuesha *negoʔ* MB, EF, FZH;

Chamicuro 7. Chamicuro *ukojka* [ukohka] MB, EF; *kohka* MB, EF (ad.);

Palikur 8. Palikur *kuku* ad. MB, EF, *nukukrin* (my) MB, (my) EF; Marawa *ukí* U;

Caribbean Island Carib *nøkəka-yɛm* MB, FZH, MZH;

Ta-Arawakan Guajiro (*ta-* = 1sg in Ta-Arawak) *tauší* HF; Paraujano *táurï* EF;

North Amazonian Resígaro *-khíigi* ad. MB; **9.** Yucuna *oku* MB, EF; **10.** Cabiyarí *aaku* MB, EF; Achagua *nukuíri* U, *kūwi* ad. U; **11.** Piapóco *nukuíri* my MB, EF, FZH, *kuí* ad. MB, EF, FZH; Curripaco *nukiri* my MB, FZH, EF, *kiiki* ad. id.; Baniwa *nū́ku* MB, A, *nukihi* MB, EF, FZH; Katapolitaní *līkeri* (*lī-* = 3sg = her/his) MB, *likeri* EF; Siusi *nukiri* MB, EF, *nuheri* GF; Sucuriu *nōkeri* MB; Adzáneni *núkiri* my MB; Jurupari *nukiri* MB, EF; Carútana *nókiri* MB, EF; Arara *nõkihi* MB, EF; Tariana *nõḥuḗ̱ri* GF, *nukhí* MB, EF (ad.); **12.** Tariana *núkhiri* MB; EF; Guarequena *nuk:o* my MB; Guinaú *nuxkú* MB, *nūku* EF; **13.** Yavitero *nō'ko* MB; Kariaí *nurey* MB; Manao *ghōko* MB; Wainumá *ghoxhoi* MB; Waraicú *ghuk* MB.

❖ Proto-Arawakan **aku(ro)* A (FZ), EM

In the cognate sets below, Payne's material is numbered from 1 to 5. As in the case of the previous cognate sets, we have eliminated consonant *h* from the reconstruction. The reconstruction of vowel **u* is founded on regular correspondences. The final ro is the feminine pronominal suffix, reconstructed in Proto-Arawakan (Maipuran) by Matteson (1972: 164) and Payne (1987: 63).

➢ **The cognate set **aku-* [**(n)aku(ru)*] my FZ/EM/(GM) in Arawak (Maipuran)** (for details of sound correspondences see Payne 1991)

Southern Arawakan 1. Bauré *-aki* EM; Ignaciano *naka* A, GMZ;

Campa Asháninka *áiro-* FZ, WM; Ashéninka *nayiro* my FZ, my EM; Ashéninka Perené *airontsi* ad. FZ (ego ♂), WM, *nayiro* EM; Nomatsiguenga *nágiro* my FZ, my EM; Matsiguenga *-agiro* FZ, MBW, EM; Nanti *pagiro* ad. EM, *obagirote* her HM;

2. Chamicuro *-ajka* [ahka] my A (ego ♀);

---

[17] The Campa languages (Asháninka, Ashéninka, Nomatsiguengua, etc.) display forms with an additional internal nasal archiphoneme (traditionally noted *N*) whose phonetic realization varies according to context. This feature belongs to the phonemic inventory of the Campa languages: it occurs after a vowel and it *"homorganically assimilates to a following obstruent"* (Payne 1981: 62, repeated in Mihas 2010: 62).

Southwestern Arawakan Piro *naxiro* GM; Iñapari *najɨró* my GM; Cuniba *jiru*; Apuriná *-akuro* ~ *nakįrú* GM;

Xingu-Paresí Yawalapití *nākiru* FZ; Mehináku *akɨ* FZ; Paresí *nakero* FZ, EM;

Palikur *nakirun* FZ, EM, *akiá* ad. FZ, EM;

Ta-Arawakan 3. Guajiro (*ta-* = 1sg in Ta-Arawak) *taʔirɯ* my A; Paraujano *táira* A; **4.** Arawak (Lokono) *takiru* HM;

Caribbean Island Carib *nakre* ~ *takɛre* my SW; **5.** Black Carib *nagoro* my SW;

North Amazonian Baniwa *nā́ko* EM, *nako* A; Curripaco *ākum* ad. FZ, MBW, EM; Baré *nako-* A, EM; Guinaú *naxkú* EM; Manao *nakueru* FZ; Mariaté *naku* 'amita' FZ.

2.3.2. Reconstruction of KOKO *in Pano-Takanan: ca. 4,700 BP*

The Pano-Takanan languages. (Map: Wikipedia).

There is a near consensus on recognizing the genetic relatedness of the Panoan and Takanan language families. Proto-Pano-Takanan has been dated back to ca. 4,700 BP by Swadesh (Kaufman 2007: 70).

The Panoan languages are now spread across the Peruvian lowlands, notably on the Ucayali River basin, the lower Urubamba River, the Upper Purús in the western part of the state of Acre (Brazil), the southwestern part of the state of Amazonas (Brazil), and finally in northern Bolivia. Our knowledge of this language group's prehistory relies on linguistics and archeology. Using glottochronology, d'Ans (1973: 364) stated that, by AD 100 (±300 years), the Panoan language family which, according to Lathrap, originated in northern Bolivia, had just differentiated into three subgroups: Proto-Preandino (Cashibo), Proto-Ucayali-Cabeceras (Proto-Ucayali-Yaminahua), and Proto-Beniano (Proto-Chacobo). Myers (1974: 135) and other archeologists assume that the arrival of Panoan speakers on the Ucayali corresponds to the appearance of ceramics of the Pacacocha tradition in the archeological record, ca. AD 300. Chacobo, Kaxariri, Pacahuara groups still live in Bolivia close to Takanan-speaking groups (Erikson 1993: map p. 46).

As far as historical documentation can trace them back (500 years or so), Takanan language groups, among which the Araona, Eparamona, Uchupiamona are known for having been established within the northern part of the angle composed by the Madre de Dios-Manu Rivers and the Beni River. This region is posited as the Proto-Takanan homeland (Wichmann *et al.* n.d.: 20).

Key (1968: 73) listed a few Takanan and Panoan *KOKO* terms in her Takanan cognate set glossed 'uncle', but did not formally reconstruct a proto-form. Girard (1971: 90) also established a cognate set and reconstructed **ku-* or **kuku* MB in Proto-Takanan. He agreed with Shell's (1965: 144, items 190 and 192) Proto-Panoan reconstructions **koka* MB and **koko* 'nephew',[18] and proposed **kuku* ~ **kuka* MB in Proto-Pano-Takanan (1971: 165). Table 3 below gives fairly complete cognate sets in Panoan and Takanan languages. Panoan languages are presented according to Amarante Ribeiro's (2006) classification.

---

[18] Speaking of her reconstructions, Shell (1965-1975: 11) used the term *'reconstructed Panoan'* (*'Pano reconstruido'*) instead of *'Proto-Panoan'*, an expression that she wanted to keep for future, more advanced reconstructive works.

➢ *Proto-Takanan* *kuku

Key (1968: 35, 37; 1992: 98, 100) and Girard (1971: 23-25) disagree on which consonant phoneme is to be reconstructed in Proto-Takanan: **k* for Girard, **x* for Key. Girard convincingly argues for **k > h, x* or *k* in Takanan languages, pointing to its relationship to Proto-Panoan **k*, and the evidence of Cavineña *k* (Girard 1971: 25). As noted in table 3, Ese Ejja and Huarayo display irregular reflexes, respectively *tsotso* and *toto* MB.[19]

Table 3. Pano-Takanan cognate sets

Proto-Pano-Takanan	***k***	***u***	***k***	***u***	**MB, EF**
Proto-Panoan	***k***	***o***	***k***	***a***	**MB, EF, FZH**
Proto-Panoan	***k***	***o***	***k***	***o***	**'nephew'**
	Group I				
Amahuaca	*k*	*o*	*k*	*á*	MB, (♀)F
	Group II				
Capanahua	*k*	*o*	*k*	*a*	U (MB), EF
Cashibo	*k*	*u*	*k*	*u*	MB, EF, FZH
Huariapano (Panobo)	*k*	*o*	*k*	*a*	U
Pano (Navarro)	*k*	*u*	*k*	*a*	MB, HF
Pano (Navarro)	*k*	*u*	*k*	*u*	'sobrina' (niece)
Shipibo-Cunibo	*k*	*o*	*k*	*a*	MB, WF
Shipibo-Cunibo	*k*	*o*	*k*	*o-*	(♀)ZS, ?(♀)BS
	Group III				
Cashinahua	*k*	*u*	*k*	*a*	MB, FZH, EF
Catuquina	*k*	*o*	*k*	*a*	MB
Isconahua	*k*	*o*	*k*	*a*	MB
Kaxariri	*k*	*u*	*k*	*u*	U, EF
Marubo	*k*	*o*	*k*	*a*	(♂)MeB, (♂)eZS, (♀)MB, (♀)ZS
Marunahua (Shell)	*k*	*ó*	*k*	*á*	U
Poyanahua	*k*	*u*	*k*	*a*	Ad. U
Shanenahua	*k*	*u*	*k*	*a*	U
Sharanahua	*k*	*o*	*k*	*a*	MB
Yaminahua	*k*	*o*	*k*	*a*	MB, EF, FZH, & ad. id. + (♂)ZS, (♀)S
Yawanahua	*k*	*u*	*k*	*a*	U
	Group IV				
Chacobo	*k*	*o*	*k*	*o*	U
Korubo	*k*	*o*	*k*	*o*	MB
Matis	*k*	*u*	*k*	*u*	MyB, FZH, EF, yZS.
Mayoruna Tabatinga (Spix: 1820)	*k*	*u*	*k*	*u*	U
May. (Matses: Erikson)	*k*	*u*	*k*	*u*	MyB, EF, yZS
May. (Matses: Fields)	*k*	*o*	*k*	*a*	MB, EF
May. (Matses Fields)	*k*	*a*	*k*	*o*	(♀)BS
Pacahuara (Armentia)	*k*	*u*	*k*	*o*	U, EF, 'sobrino'
Pacahuara (d'Orbigny)	*k*	*o*	*k*	*o*	U
Proto-Takanan	***k***	***u***	***k***	***u***	**MB**
Araona	*h*	*u*	*h*	*u*	U
Cavineña	*k*	*ʊ*	*k*	*ʊ*	MB (ad.)
Ese Ejja (*ts* irregular)	*ts*	*o*	*ts*	*o*	MB
Huarayo (*t* irregular)	*t*	*o*	*t*	*o*	MB
Reyesano	*h*	*υ*	*h*	*υ*	U
Tacana	*h*	*u*	*h*	*u*	MB

➢ *Proto-Panoan* *koka *and* *koko

Shell (1965-1975: 144, items 190 and 192), followed by Girard (1971: 165), reconstructs two roots **koka* and **koko* in Panoan on the basis of regular sound correspondences.

Girard (1971: 165) glossed **koka* 'maternal uncle' upon the overwhelming evidence of the reflexes referring to this relationship in groups I, II and III.[20] The same (1971: 165) glossed **koko* 'nephew' upon the evidence presented by Shell (2008: 144), namely Pano (Navarro) *kuku* 'sobrina' [niece], Shipibo-Cunibo *koko* (♀)ZS, 'sobrino', 'hijo de la hermana de una mujer' [nephew], Cashibo *koko* EF, Chacobo *koko* 'tio' [uncle], Mayoruna *kuku* 'tio [uncle], suegro [EF], sobrino

---

[19] In Ese Ejja, a few words display an apparent evolution **k > x*, e.g. **kunu* > Takana *hunu* 'liana,' Cavineña *kono* 'liana,' Ese Ejja *xono* 'liana;' **kuři* > Takana *huri* 'ocelot,' Cavineña *ho-kori* 'badger,' Ese Ejja *xowi* 'ocelot,' etc. (Girard 1971: 90).

[20] *Koka* MB was also elicited in group IV (Matses) by Fields in 1970 (Erikson 1986: 192, 201), but has never been recorded since.

[nephew] (details in table 3),' Pacahuara *koko* 'tio, suegro, sobrino'. To these must been added Matis *kuku* which refers to MB, EF and yZS. But there are also a few *koka* forms meaning both 'nephew' and MB, EF, such as Marubo *koka* MeB, (♂)eZS, (♀)ZS, Yaminahua *koka* ref. MB, EF, FZH & ad. MB, EF, FZH, (♂)ZS, (♀)S. As a matter of fact, the opposition between *koka* MB, EF on the one hand, and *kuku* 'nephew' on the other hand is not so clear-cut, although the presence of reflexes of both forms in all the groups including more than one language indicates, without much possible doubt, that they have their origin in Proto-Panoan. It is important to note that all the languages of group IV, except Matses, display *kuku* ~ *koko* forms while most languages from the other groups display *koka* forms.

We must also contemplate the hypothesis that *kuku* ~ *koko* forms were borrowed from other language families, maybe from Arawakan. Table 4 recapitulates the various *KOKO* terms found in Arawakan (vocative) and Panoan, suggesting possible mutual borrowings. Southwestern Arawak languages, including Piro, Cuniba, Kanamaré, Cushichineri[21] appear as the best candidates in this respect. The problem is that the northern Panoan languages of group IV (Mayoruna, Korubo, Matis and Matses)[22], all having *koko* or *kuku* forms, are separated from the Southwestern Arawak languages by the main body of Panoan languages, all having *koka* forms. This is not the case for the Bolivian languages of group IV (Chacobo and Pacahuara) which are neighbors with the Southwestern Arawak language groups displaying *kuku* forms. These Bolivian Panoans are also neighbors with the Takanan groups which, as we know, also display *huhu* or *kuku* forms. The similarity of *KOKO* forms in all these languages could have resulted from ancient areal diffusion.

Table 4. Comparison of Arawakan and Panoan *KOKO* forms

Panoan *KOKO* forms		**Arawakan *KOKO* address forms**	
Panoan *kuku* ~ *koko* forms		**Southwestern Arawak *kuku* ~ *koko* forms**	
Cashibo	*kuku* MB, EF	Piro	*koko* MB, HF, WFB
Mayoruna (Matses)	*kuku* MB, EF	Cushichineri	*koko* U
Matis	*kuku* MB, EF	Cuniba	*kuku* MB, EF
Chacobo	*koko* U	Kanamaré	*ghughu* MB
Panoan *koka* ~ *kuka* forms		**Southwestern Arawak *koka* forms**	
Amahuaca	*kuka* MB	Mashko	*kokoa* U
Shipibo-Cunibo	*koka* MB, EF	Sirineiri	*kokoa* U
Capanahua	*koka* MB, EF	**Chamicuro**	*kohka* MB, EF < Panoan?
Panobo	*koka* U	**Campa forms**	
Isconahua	*koka* MB	Asháninca	*kokí* my (♂)MB, EF
Cashinahua	*kuka* MB, FZH, WF	Ashéninca	*koko-ini* ad. (♀)U, (♀)EF
Catuquina	*koka* MB	Matsiguenga	*koki* MB, EF
Marubo	*koka* MB	Nomatsiguenga	*koki* ad. (♂)MB, EF
Sharanahua	*koka* MB	**Other Southwestern Arawak forms**	
Shanenahua	*kuka* U	Apuriná	*kįkį* 'man'
Yawanahua	*kuka* U	**Upper Rio Negro forms *koko* forms**	
Yaminahua	*koka* MB, EF	Resígaro	*-khíigi* MB
Poyanawa	*kuka* U	Achagua	*kúúwi* ~ *kúwi* MB
Marunahua	*kókó* U	Piapoco	*kuí* MB, EF

---

[21] Spoken or formerly spoken in the north of the Madre de Dios River, and along the Purús and its tributaries (southeastern Peru and Brazil).

[22] Spoken along the Javari River and its tributaries (along the border between northeastern Peru and Brazil).

Setebo	*kuka* U, HF	Curripaco	*kiiki* ad. MB

Denotata have been simplified in this table.

- ***Proto-Pano-Takanan* *kuku/a**

Girard (1971: 165) reconstructed **kuku/a* MB in Proto-Pano-Takanan on the basis of regular sound correspondences between Proto-Panoan and Proto-Takanan (1971: 158-9), leaving open the nature of the second vowel (*u* or *a*).

2.3.3. Reconstruction of KOKO *in Proto-Arawan: 1,764 BP (ASJP)*

The Arawan language family is composed of 5 languages: Sorowahá, Madi, Paumarí, spoken in the state of Amazonas (Brazil) along the Purús River and its tributaries and the Juruá River, Kulina-Dení, spoken along the upper Purús and the Juruá River, and finally Arawá which is extinct and, for the most part, undocumented. The four living languages and their dialects display *KOKO* forms, referring to MB and EF (table 5). Note that de Créqui-Montfort and Rivet (1922: 175) mentioned the existence of Paumarí *kukŭ* EM and Jamamadi *koko* S. We have not been able to confirm or infirm the existence of these meanings in more recent data.

Table 5. Arawan comparative data

Proto-Arawan	***k o k o***	**MB, (EF)**		
Sorowahá	*k o k o*	MB, EF		
Madi				
Jarawara (dial.)	*k o k o*	EF (ad.)		
Jamamadi (dial.)	*k o k o*	MB, EF, S		
Banawá (dial.)	*k o k o*	MB, EF		
Paumarí	*k o k o*	MB, EF	*k u k ŭ*	EM
Kulina-Dení				
Kulína	*k o k ó*	MB, EF		
Dení	*k o k o*	MB, EF		
Madihá-Zuwiha (dial.)	*k ó k o*	HF		
Madihá-Kurina (dial.)	*k u k u*	U, EF		

Language names are in bold.

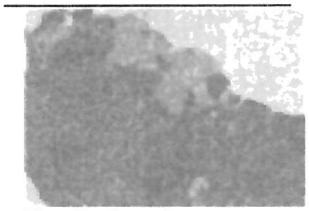
Dixon (2004b: 48, 61) reconstructs **koko* MB, EF in Proto-Arawan (**k* > *k* in all languages and all environments, and **o* > *o* in all languages and all environments), but considers this form as an Amazonian areal feature, and concludes, as we already mentioned (2004b: 17), that it *"could have been borrowed from anyone of a number of sources in Proto-Arawan, or else into individual languages."* He also claims that Arawan has no genetic link with Arawakan (Maipuran) whatsoever. A likely location for Proto-Arawan could be between the lower Purús and the Juruá River, where the vast majority of Arawan languages (including Arawan until its extinction) are spoken. No date is suggested for Proto-Arawan, except by ASJP: 1764 BP.

2.3.4. Reconstruction of KOKO *in Proto-Cariban: ca. 3,700 BP*

Proto-Cariban is estimated back to ca. 3,700 BP (Kaufman 2007: 75); its homeland is posited in Venezuelan Guiana (Villalón 1991; Heckenberger 2002: 103). De Goeje (1909: 30) first published a list of Cariban *KOKO* forms, on the basis of which he later proposed **koko* U (1946: #

233, p. 58)[23]. The meaning of the reconstructed masculine form surely includes MB, but also possibly EF and GF. This list has now been considerably augmented, and one can remark that there is a basis for postulating also a feminine proto-form *kuku* ~ *koko* GM, FZ, EM,[24] reconstructed as **kuku* GM in Proto-Taranoan by Meira (1998: 172, # 118), on the basis of **k* > *k* in all languages and all environments, and **u* > *u* in all languages and all environments.

The date of the Proto-Taranoan split calculated by Meira (1998: 159-160) is between 500 and 900 years ago. Our "inspectional reconstruction"[25] $^{\#}$*kuku* ~ *koko* GM in Cariban is founded on the occurrence of *koko* or *kuku* GM in most of the branches composing the Cariban family, as defined in Gildea's classification (2012: 445), i.e., Venezuelan, Nahukwa, Guianan (**kuku* in Proto-Taranoan), as well as in some unclassified languages (see the comparative data below). Similar occurrences of GM forms are also massively found in Chibchan languages, in Mapudungu and in North and Meso-American language families (see Appendix and tables 11, 12). *KOKO* terms in Cariban are address terms. Reference terms referring to the same kin types can be reconstructed (notes 23 and 24).


The Cariban languages. (Map: Wikipedia).

Cariban comparative data. Classification of languages adapted from Gildea (2012).

$^{\#}$*koko* MB, (GF, EF), $^{\#}$*kuku* GM, (FZ, EM);

Parukotoan Branch: Hixkaryana (Parucoto) *owhoko* U, EF;

Pekodian Branch: Bakairí *kogo* MB, FZH; Arara (Pará) *koko* ~ *kɔkɔ* MB, FB, FBS, FZS;

Venezuelan Branch: Makushi *kooko*, *okoko* GM, *koko* MB; Pemón *koʔwai* ad. GM; Taurepang (Pemón dial.) *koʔway* ~ *kokó* (my) GM; Arekuna (Pemón dial.) *kɔkɔi* GM; Akawaio *'aʔai* M; Panare *koko* FF, (♂)FBS, (♀)MZD, (♂)eB, *kokon* yB, young C; Tamanaco a *koko* form?;

Nahukwa Branch: Kuikúru *kokojó* ref. GM, other fem. relatives, *o'o* ad. GM; Kalapalo *okojo*, *oó* GM, -*kuẽgü* GGP;

Guianan Branch: Carib (Kari'nja, Galibi) *molekóko* boy, *kah-tobo* MB, WF, (♂)FZH, MBS, FZS; Carib (Cachama) *káx-topo* GF, MB, EF; Carib (Oiapoque) *ka-tobô* MB, MBS, (♀)FZS; Carib (Maroni River) *kax-tobo* MB, (♂)MBS, (♀)FZS, *kooki* B, MBS, S, SS, etc.; Ye'kwana (Makiritare) *kóóko* ~ *kóko* ad. GF, FFZS etc.; Wayana (Roucouyenne) *ku-ni* GM, old woman, *kɔ̃nko* FF, MB, EF, (♀)HB; *akon* eB, *kono* WB, HZ; *Taranoan group:* ***kuku GM, (FZ, EM);** Pianacoto *ku-ni* GM; Akuriyo (Triometesem) *kuku-ni-komo* GM; Trio *kuku* ad. & ref. (♂)GM, (♂)FZ, WM, (♂)MBW, (♀)GM, (♀)FZ, HM, (♀)MBW, *konóka* ref. (♂)FZH, WB, ZH, etc., *koko* ref. & ad. (♀)FZD (rarely), (♀)MBD, (♀)BW, HZ; Carijona *kuuku* GM, FZ; Hianácoto-Umáua *kúúku* GM;

Residue: Apalaí *e-o-ø* 'my MB'; *Yukpa group:* Japréria *yuvan-koko* A; *Unclassified:* Sapara *kuu-nutú* GM; Apiaka *koko* U; Pimenteira *kuckú* Oheim (MB), boy; Paravilhana *gocko* MB, *tamuy gocko* GF.

---

[23] Another Cariban (reference) term refers to MB and EF, probably from Proto-Cariban **j-awo-* (de Goeje 1946: 58).

[24] Another Cariban term refers to GM and FZ, probably from Proto-Cariban *-*noti* (de Goeje 1946: 57).

[25] The symbol $^{\#}$ is the convention that Blench adopted from the Niger-Congo volume edited by Bendor-Samuel (1989), to distinguish "*reconstructions established by regular sound correspondences* [marked with an asterisk *, AM & PB] *from those derived by quick inspection of cognates … The effect of this is to translate the starred forms of various writers to hache* '#'." (Blench 2008: 204).

2.3.5. Reconstruction of KOKO *in Proto-Guahiboan: ca. 2,300 BP*

The Guahiboan languages. (Map: Wikipedia).

Kaufman (2007: 65) gives 23 centuries for Proto-Guahiboan, whose homeland is posited by Wichmann *et al.* (n.d.: 19) on the northeastern side of the Colombian Andes. Christian & Matteson's article on Proto-Guahiboan (1972: 150-51) is rather scanty and based on a comparison between 3 languages: Guahibo, Cuiva and Guayabero. Unfortunately, we have not been able to access Keels' more recent study (1986) on Proto-Guahiboan, which is focused on the same languages. Correspondences – existing in almost all environments – given in table 6 are from the first authors. Correspondences in Macaguán are deducted from available comparative material extracted from Buenaventura (1993). We do not have the term for MB in Guayabero but the sound correspondences in this language, deducted from Keels (1985), are probably **x > x, *k > k, *u > u*. Thus we just venture a reconstruction of **axu-jo* MB, EF, FZH and **akwe* FM, MM in Proto-Guahiboan. Note that *-jo* is a diminutive suffix.

Table 6. Guahiboan comparative data

Proto-Guahiboan	***a***	***x***	***u***	***j***	***o***	**MB, EF, FZH**	***a***	***k***	***w***	***e***	**GM**
Cuiva (Hiwi) **a > a, *x > x, *u > u, *j > j , *o > o, *k > k, *w > w, *e > e*	*a*	*x*	*u*	*y*	*o*	MB, WF	*a*	*k*	*w*	*e*	GM
Sikuani **a > a, *x > x, *u > u, *j > j , *o > o, *k > k, *w > w, *e > e*	*á*	*x*	*u*	*y*	*o*	MB, EF, FZH	*á*	*k*	*w*	*e*	GM
Guahibo (Rivet 1948)	*a*	*x*	*ui*	*y*	*o*	EF					
Guahibo (Rivet 1948)	*a*	*k*	*ú*	*y*	*o*	EF					GM
Guahibo (Rivet 1948)	*a*	*k*	*u*	*é̲*		U					
Macaguán **x > k, *k < k, *a > a, *u > u, *e > e*	*a*	*k*	*ú*	*i*		EF (ad.)	*a*	*k*	*u*	*e*	GM

2.3.6. Reconstruction of KOKO *in Proto-Chibchan: ca. 6,700 BP*

The Chibchan languages. (Map: Wikipedia).

Constenla Umaña (2009: 209) gives 6,550 BP (6682 BP in Constenla Umaña 2012: 419) for the split between Pech (Paya) and Core Chibchan, and suggests a Proto-Chibchan homeland in Costa Rica and Panama (Constenla Umaña 2009: 209). This date matches the 7,000 years obtained through genetic studies of populations from this language group (Hoopes & Fonseca 2003: 61). Constenla Umaña (1981: 381, 399-400) reconstructed the following items in Proto-Chibchan:

1. **káka* F (cognate set numbered 1 to 3 in table 7), with **k* having sound correspondences in Cabécar *k* (all positions), Teraba k^h (word initial position), *k* elsewhere, Dorasque *g* (intervocalic position), *k* elsewhere, and **a* having sound correspondences in Cabécar *a* (all environments), Terraba *o* (all environments except in contact with laryngeal or nasalized), Dorasque *a* (all environments).

2. **gAkA* EM[26] (cognate set numbered 4 to 7 in table 7), based on sound correspondences: **g* > Cabécar *h* before *i* and *u*, *ĵ* elsewhere, Bribri *ø* before *i* and *u*, *ĵ* elsewhere, Cuna *s* in all positions, Cágaba *ž* before *ɪ* (one etymology), *g* elsewhere; **k* > Cabécar *k* in all positions, Bribri *k* in all positions, Cuna *k* in all positions, Cágaba *g* in intervocalic position; **a* > Cabécar *a* in all positions, Bribri *a* the same, Cuna *a* the same, Cágaba *a* the same. As for the final vowel *-i* of Cágaba *gagi* EM, it is explained by Constenla Umaña (1981: 362) as a nominal suffix very frequent in this language.

Table 7. Chibchan comparative data (from Constenla Umaña 1981 and additional material)

Proto-Chibchan		*	***k***	***á***	***k***	***a***	**F**		*	***g***	***a***	***k***	***a***	**EM**
Pech														
Pech (Paya)											*u*	'	*a*	FZ, EM
Core Chibchan														
Isthmic			***k***	***a***	***k***	***a***	**F**							
Cabécar	**1**		*k*	*a*	*g*	*a*	F							
Cabécar								**4**		*ǰ*	*â*	*k*		EM
Cabécar (Estrella)	**1**		*k*	*a*	*k*	*a*	F							
Cabécar (Čiripo)	**1**		*k*	*a*	*k*	*a*	F							
Cabécar (Tukurrike)	**1**		*k*	*a*			F							
Bribri								**5**		*ǰ*	*á*	*k*		EM
Teribe	**2**		*k*	*o*	*k*		F							
Teraba	**2**		*kʰ*	*ó*	*k*		F							
Dorasque (Chángena)	**3**		*k*	*a*	*g*	*a*	F							
Cuna								**6**	-	*s*	*a*	*k*	*a*	EF
Magdalenic														
Cogui (Cágaba)								**7**	-	*g*	*a*	*g*	*i*	EM
Guamaca (Malayo, Damana)									-	*gw*	*a*	*g*	*i*	EM
Bíntukua (Ica)			*k*	*á*	*k*	*ə*	F		-	*gw*	*a*	*t*	*i*	A
Atanques (Cancuama)			*k*	*á*	*k*	*e*	F			*gw*	*a*	*s*	*i*	A
Central Tunebo (U'wa)										*k*	*a*	*g*	*i*	FZD, MZD
Central Tunebo (U'wa)										*gʷ*	*a*	*k*	*a*	(♀)EF, (♀)DH
Not classified														
Nutabé										*gu*	*a*	*k*	*u*	old woman

The additional list below presents more data than what was used by Constenla Umaña in 1981.

Additional Chibchan comparative data:

a) Proto-Chibchan *#kVkV-* MB, EF, GF;

---

[26] Constenla Umaña (1981: 399) reconstructed **gAkA* EM, with *A* indicating that the proto-form had either **a* or **ə*. The contrast between these vowels is no longer recognized by the author (2012: 404-5). Consequently, we wrote **gaka* in table 6.

Pech Pech (Paya) *uku* MB, EF; *oká* B;

Core Chibchan

Isthmic Cabécar *këgölö* [kɪgɷlɷ] MB; Cabécar *kíkíwa* 'anciano(a);' Cabécar (Estrella) *keke* 'señor'[27]; Cabécar (Čiripo) *keke* 'señor'; Bribri *akëgëla* [akígɪla] 'anciano'; Boruca *kakat* B, *sugu* 'anciano'; Teribe *kʰòkí* GF, *kʰɛ́gɛ́* MB, *kʰɛ́gɛ́* EF; Teraba *kega* U, *kegi* EF; Dorasque (Chumulu) *wóka* GF; Dorasque (Gualaca) *óka* GF; ***Magdalenic*** Atanques (Cancuama) *kuku* U, *sukwi* GM; Bíntukua (Ica) *tegue* U; Damana (Malayo) -*kugu* U, GF; Cogui (Cágaba) -*kukwi* A;

b) Proto-Magdalenic: ~ 5,200 BP (Constenla Umaña 2012) #*kaka* GM; Bíntukua (Ica) -*zaga* GM; Damana (Malayo) -*takə* GM; Cogui (Cágaba) *kaka* GM; Tunebo (U'wa) *kaka* GM; Muisca (Chibcha) *kaka* GM; Tunebo Sínsiga of Cobugón *kaka* GM.

2.3.7. Reconstruction of KOKO *in Proto-Lenca-Misumalpan: ca. 7,200 BP*

The Lenca family is composed of 2 languages. One of them is spoken in Honduras, the other one in El Salvador. The Misumalpan language family includes 4 languages: Miskito (Nicaragua), Matagalpa (Nicaragua), Cacaopera (El Salvador) and Sumo (numerous dialects in Nicaragua). A number of authors have postulated a genetic relationship between these 2 families, notably Constenla Umaña (2002: 189). In an attempt to establish this idea, Constenla Umaña established a list of 92 joint Lenca-Misumalpan sets (2002: 191-193), among which one finds #7: 'anciano(a),' thus Cacaopera *kukus*, Sumo *kukuŋ*, Miskito *kuka*, Lenca-Salvador *koko*. The author's list can be extended, adding similar words from other Misumalpan and Lencan languages or dialects (table 8). Constenla Umaña (2002: 193-196) also observed a significant number of phonetic correspondences between Misumalpan languages and Lencan languages, notably Misumalpan **u* = Lenca **o*, or Misumalpan **k* = Lenca **k*. Koontz-Garboden and Francez (2009: 9) reconstructed the 1sg possessive suffix **-ki* in Proto-Misumalpan.

Constenla Umaña's glottochronological study, developed in the same article (2002: 197-202), and based on 120 Lenca-Misumalpan sets, furnishes dates for several language splits. The Lenca-Misumalpan split apparently happened ca. 7,200 BP. Within Misumalpan, the split between Proto-Miskito and the other Misumalpan languages occurred ca. 5,800 BP. The split between Matagalpan (Matagalpa and Cacaopera) and Sumo-Ulwa languages occurred ca. 5,300 BP. Other conclusions concerning further subdivisions were also reached, notably the date of the split of the Lenca dialects ca. 2,300 BP.

Table 8. Lenca-Misumalpan comparative data

Proto-Lenca-Misumalpan	***k**	***V**	***k**	***V**				**GF, U, EF**
Lenca	***k**	***o**	***k**	***o**				
Lenca (Guajiquiro - Honduras)	*k*	*o*	*g*	*o*				EF
Lenca (Chilanga - El Salvador)	*k*	*o*	*k*	*o*				'anciano'
Lenca (Chilanga - El Salvador)	*k*	*o*	*h*					U, eB
Misumalpan	***k**	***ū**	***k**	***u**	-	***k**	***i**	my GF, (U, cousin)
Miskito	*k*	*ū*	*k*	*a*				GM
Miskito	*k*	*ú*	*k*	*i*		*k*	*i*	(my) GM
Matagalpa	*k*	*u*	*k*	*u*		*k*	*e*	(my) U
Cacaopera	*k*	*u*	*k*	*ú*	*l*			GF

---

[27] Constenla Umaña (1981: 403-4) reconstructed 'old' as **A'kíkí.*

Cacaopera	*k*	*u*	*k*	*ú*	*s*			GM
Karawala Ulwa (Green)	*k*	*ū*	*k*	*u*	*ŋ*	*k*	*i*	(my) GF
Pearl Lagoon Ulwa	*k*	*u*	*k*	*u*	*ŋ*			GF
Sumo (Lehmann)	*k*	*u*	*k*	*ō*		*k*	*ë*	(my) cousin
Sumo (Lehmann)	*k*	*ū*	*k*	*ŭ*	*ŋ*	*k*	*ë*	(my) GF
Sumo (Von Houwald)	*k*	*u*	*k*	*u*	*ŋ*	*k*	*i*	my GF
Panamahka Sumo	*k*	*ū*	*k*	*u*	*ŋ*	*k*	*i*	my GF
Tuahka (Twaka)	*k*	*o*	*k*	*o*				GF

2.4. Other KOKO *cognate sets in South American linguistic families*

2.4.1 KOKO *in Quechuan and Aymaran: at least 1,500 BP*

Quechuan and Aymaran are the two South American indigenous language families most widely spoken today in the middle Andean area. Quechuan "dialects" or "languages" are spoken by some 8 million people along the Andean cordilleras, from Southern Colombia to Northeastern Argentina. The Quechuan language family is divided into 2 branches: Quechua I (or B), henceforth Q I, comprising dialects occupying a continuous area in the central Peruvian Highlands, and Quechua II (or A), henceforth Q II, comprising the remaining Quechuan varieties, situated both to the north and the south of the central Peruvian zone, from Ecuador to Bolivia and Argentina (Adelaar and Muysken 2004: 185-186, Adelaar 2012: 578-580). The Aymaran family also comprises 2 branches: a northern branch with 2 languages, Jaqaru and Kawki, spoken by a few thousand people (Kawki by just a few) in the Peruvian province of Yauyos (southeast of Lima), and a southern branch, with more than 2 million speakers, straddling Bolivia, Chile and Peru (Adelaar 2012: 577-8). The distribution of these two language families results from their pre-Colombian, but also from their colonial expansions. Soon after the Inca Empire was defeated, Quechua and Aymara were promoted, along with Puquina, as "general languages" for colonial administrative and religious purposes (Adelaar and Muysken 2004: 167). So they continued beeing spoken at the expense of other indigenous Andean languages, many of which were still spoken at the time of European contact (Adelaar and Muyskens 2004: map 3, p. 166).

A consensus has been reached pointing to Central Peru as the homeland of both Proto-Aymaran and Proto-Quechuan, and to date their respective initial linguistic divergence before 500-600 AD (Adelaar and Muysken 2004: 181, Heggarty 2008: 39, 46-47 and 52)[28]. The Aymaran language seems to have been the first language to expand, southwards towards Cuzco and Southern Peru, and possibly northwards towards Ancash (Heggarty 2008: 39-41, 48-49), covering a region roughly corresponding to Huari's territorial influence. The development of Aymaran on the Altiplano seems to have happened much later, probably after 1000 AD (Heggarty 2008: 50). Quechua's first linguistic divergence was apparently limited to a region from the north of Ancash to Huancayo in the south (Heggarty 2008: 50 and fig. 6). This is approximatively where the Central Quechuan (Q I) dialects are now spoken. The general expansion of Quechuan languages, towards Ecuador and the Cuzco region, was initiated long after this first linguistic spread, but *"still a few centuries before the Inca conquest"* (Heggarty 2008: 50). It is debated whether the Late

---

[28] Adelaar (2012: 588) says that *"More recently a different scenario has emerged as the expansion of Quechuan was attributed to the centralized state of Huari."* According to this scenario, the region of Ayacucho, where the capital of Huari was situated, could have been either the homeland of Quechuan as a whole, or of Quechua II alone (Adelaar 2012: 588, references included).

Intermediate Period (1000-1476) corresponds to this language expansion phase, or if it is the Middle Horizon (600-1000: Huari culture), that was primarily equated with Aymaran. During this expansion phase *"it is unclear how significant a population movement was involved"* (Heggarty 2008: 51), but it is generally admitted that, notably in the northern region, *"this form of Quechua spread principally as a* lingua franca *for trade"* (Heggarty 2008: 51). Apparently, the Incas were responsible for taking Quechua further south into Southern Bolivia, leap-frogging the Aymara-speaking regions of the Altiplano (Heggarty 2008: 52).

Quechuan and Aymaran have much in common linguistically and their genetic relationship has been regularly debated. Their relationship is a matter of millennia, but it is now questioned if these two families have any genetic relationship at all. Most linguistic convergences have been attributed to heavy language contact (Heggarty 2008: 43-46).

A number of Quechuan and Aymaran languages refer or referred in the past to the maternal uncle and the father-in-law, using the word *kaka*.

➢ Kaka *in Quechuan languages*

The term *kaka* has been recorded in Quechuan languages during the 16th and 17th centuries in various regions of Peru.

a. On the central coast around Lima:

In his *Lexicon* and *Grammatica*, both published in 1560, the Dominican Santo Tomás reported the forms '*cacay* [kakaj] *abuelo de mi muger'* (WGF) (1560a: 2 v.), '*caca* [kaka] *suegro, padre de la mujer'* (WF) (1560a: 98), '*caca* [kaka] *tio, hermano de madre'* (MB), *'caca* [kaka] *tio, hermano de abuelo, hermano de abuela'* (U, GFB, GMB) (1560a: 100), *'caca* [kaka] *tio hermano de madre'* (MB), '*cacay* [kakaj] *abuelo de mi muger'* (WGF) (1560a: 111 v.), *cáca* [kaka] *'dize el sobrino al tio'* (U) (1560b: 8). It is alleged that the language described by Santo Tomás was a now extinct variety of Quechua, spoken on the coast around Lima and generally considered as belonging to the Q II branch of Quechuan[29]. Adelaar and Muysken (2004: 182, 191) and other authors believe that this form of Quechua was likely identical to the language of the Inca's administration (Inca general language). So the word *kaka* not only belonged to the lexicon of the coastal dialect but was understood everywhere across the Inca empire.

b. Around Lima and in the Central Peruvian sierra:

Among the native religious practices ("superstitions") described in his book *Extirpación de la Idolatria del Piru*, the Jesuit father Arriaga (1621: chap. 6, p. 33-4) mentioned a ceremony where four or five year old children had their hair cut for the first time, and had their names changed. At this occasion, the relatives and specially the *"cacas* [kakas] *y massas"* were invited.[30] Arriaga is not specific about the places where such ceremonies were held, but he observed that the locks of hair that were cut and considered as sacred objects, were differently called according to the region where the event occurred, either in the *sierra* or in the *llanos* (the lowlands near Lima). In his first chapter, Arriaga mentioned the places that had been already visited by the ecclesiastical personnel in charge of the extirpation of idolatry, covering parts of the Central Peruvian Sierra as well as the Coastal Lowlands around Lima[31]. This statement, as well as other information collected from early colonial documents and pertaining to different parts of Peru, make us believe that such haircuts were common practice throughout Peru[32] in the pre-colonial period.

---

[29] Cerrón-Palomino (1990: 340) notes that the form of Quechua found in Santo Tomás' *Lexicon* and *Grammatica* has much to do grammatically with the southern dialects, and lexically with the central and northern dialects.

[30] *Mas(s)a*, in 16th century Quechua from Ayacucho and central Peruvian dialects, denotes the affinal relationships of brother-in-law and DH. About the *mas(s)a-caca* relationship, see Zuidema (1977: 259, 261-262) and Webster (1977: 39).

Moreover, although carefully differentiating between the various regional terms designating the hairlocks, Arriaga does not mention any terminological variation for mother's brother and brother-in-law. *Caca* [kaka] *"tio hermano de madre"* and *masa "cuñado, pariente de afinidad"* are among the word listed in the index of his book (1621: *Indice de algunos vocablos*). We also find these words in the "edict against idolatry" published in the same book (1621: 129-133). This edict, covering the Middle Andean and coastal parts of the vice-kingdom of Peru, contains a list of questions that ecclesiastical visitors were supposed to ask the inhabitants of the towns they visited, in order to unveil native superstitions. Under item 14, it was notably asked to the natives to denounce people who they knew were performing haircut ceremonies, to which were invited the *"tios"* and the *"cuñados"* – called *"caca"* and *"masa",* respectively. Put together, these pieces of information allow us to infer that *"caca"* [kaka] was the general term for MB in the regions already visited, and mentioned by Arriaga in his first chapter, as well as those not yet visited. Arriaga's "sierra" regions are those where many Quechuan languages of the Q I branch are spoken today.

c. *In the Cuzco region:*

Numerous colonial documents, glossaries, and grammars, dated from the second half of the 16th century or shortly after, mention the existence of the word *caca* [kaka] MB in the Cuzco dialect of Quechua (Yaya 2008a, 2008b). Three publications from this period deserve to be mentioned in our paper, because of the invaluable linguistic material they contain, related to kinship and to kinship terms. The first one is an anonymous *Arte y Vocabulario* printed in 1586 by Antonio Ricardo, whose author is probably Alonso de Barzana (Zuidema 1977: 242), in which we find – notably in the *Annotaciones* – the word *caca* [kaka] or *cacay* [kakaj = 1st person possessive], glossed: *'(cacay) hijo de mi tio hermano de mi madre'* (MBS), *'(cacay) tio del varon o dela mujer, hermana de su madre'* (MB), '*(caca) dize el consuegro padre del hijo a su consuegro, y el a el, cacay'* (♂SWF), *caca dize la consuegra madre del hijo a su consuegro* (♀SWF), *'(catay) dize el suegro al yerno, y el a su suegro caca'* (WF), *'(cacay) dize el yerno a su cunñado hermano de su muger'* (WB). The second and third documents are the *Gramatica* and *Vocabulario*, written by Gonçález Holguín and published in 1607 and 1608, respectively, where we find *caca* [kaka] glossed: *'tio hermano de madre'* (MB) (1607: 96 v.; 1608: 35), *'los consuegros*

---

[31] Arriaga (1621: Chap. 1, notably p. 7) stresses the fact that the rites described in his book had been directly observed by him, or observed by other visitors of as much authority as himself, in the provinces of Huarochiri, Yauyos, Jauja, Tarma, Huaylas, Chinchaycocha (Junín), in the districts of Andajes (province of Oyón) and Checras (province of Huaura), in the city of Huanuco, etc., all situated in the archbishopric of Lima. In many of the regions cited, dialects of the Q I branch of Quechuan are spoken.

[32] Similar ceremonies stressing the role of the maternal uncle were held throughout Peru at the time of the Spanish conquest. This haircut is sometimes performed by the "closest uncle" on one-year-old infants, and called *rutuchiku*. For Chinchaycocha (department of Junín), see Duviols (1974: 277-8). For Cuzco, see Yaya (2008a: 71, and 2008b: 208).

varones entre si' (♂CEF), *'caca, el suegro del yerno que es el padre de su muger* (WF), *que al suegro y a su padre le llaman el hierno y su hermanos y primos'* (WF, WFF) (1607: 98), *'caca, cunado del, que el cassado llama a todos hermanos y primos de su muger cacay'* (WB, W's ♂cousins) (1607: 98, 98v), or *'cacay, dizen tambien los hermanos del varõ casado como el a sus cuñados varones, hermanos de su cuñada'* (WB, ♂BWB) (1607: 98v). The Quechuan dialects of the Cuzco region belong to the Q IIc branch of Quechua[33].

To our knowledge *kaka* has been preserved in modern Quechua of the Cuzco region: in Pitumarca (Milicic 2011), and by modern Q'ero who use *kakay* to refer to MB, MBS, MMBS (Webster 1977: 30). It is also used referring to MB in the Quechua spoken in Ayacucho, and also referring to MB, EF in the Quechua spoken in the Apurimac region.[34]

Table 9. Quechuan comparative data.

Proto-Quechuan	***k**	***a**	***k**	***a**	***y**	**(my) MB, WF, WB**
Santo Tomás (1560): Inca general language, Costal Quechua: Q IIb?	*k*	*a*	*k*	*a*	*y*	(my) MB, WF, WGF etc.
Arriaga (1621): Region of Lima and Central Sierra: Q II , Q I	*k*	*a*	*k*	*a*		MB
Cuzco (1552, 1570, 1586, 1607): Q IIc	*k*	*a*	*k*	*a*	*y*	(my) MB, WF, WB, MBS
Huarochiri (ca. 1608): Q IIc?	*k*	*a*	*k*	*a*		MB, EB
Q'ero (Modern): Q IIc	*k*	*a*	*k*	*a*	*y*	(my) MB, MBS
Pitumarca (Modern): Q IIc	*k*	*a*	*k*	*a*		MB
Ayacucho (Modern): Q IIc	*k*	*a*	*k*	*a*		MB
Apurimac (Modern): Q IIc	*k*	*a*	*k*	*a*		MB, EF

Proto-Quechua had a 3-vowel system **a*, **i*, **u* preserved in present-day languages with a wide range of allophonic realizations. Consonant **k* has been preserved in all languages (Adelaar and Muysken 2004: 194-198). The cognate set *kaka* reflects Proto-Quechuan **k* > *k* and **a* > *a*. Thus we can propose **kaka* MB, WB in Proto-Quechuan with some confidence, all the more so as *kaka* was in use 500 years ago in all Quechuan branches.

➢ Kaka *in Aymaran languages*

Hardman (Pyle 1981: 89, editor's [i.e. Hardman's] note) says that Aymara also referred in the past to MB as *kaka*. But this word is mentioned neither in the *Arte y gramatica muy copiosa de la lengua Aymara*, nor in the *Vocabulario de la lengua Aymara* both written by Ludovico Bertonio and published in 1603 and 1612, respectively. What we find in the *Vocabulario* (1612, vol. 2: 191) is *"Lari: tio hermano de la madre y casi a todos los varones parientes de parte de madre llaman lari"* ("MB and all the masculine parents on the mother's side are called *lari*"). This word is still used today in modern Aymara of Southern Peru at least, with the meaning WB, and rarely, if ever, with the meaning MB (Collins 1981: 218, 234-5).

---

[33] The authors want to thank Pierre-Luc Abramson, professeur honoraire des universités (University of Perpignan, France), researcher at the CRHISM (Centre de recherches historiques sur les sociétés méditerranéennes), and César Itier, maître de conférences at INALCO (Institut des langues et civilisations orientales) for their assistance in the translation of the Golden Age Castilian transcriptions of Quechuan and Aymaran kinship terms.

[34] Most Quechuan dialects today refer to both maternal and paternal uncles as *tiyu* (from Spanish *tio*). Hardman (1982: 148) states that *kaka* was replaced by *tiyu* because of *"interlingual taboos and the preoccupations on the part of the priests regarding what they imagined to be incest"*, probably the fact that *kaka* referred to both MB and EF.

In contrast, modern Jaqaru and Kawki (Hardman 1983: 204, 206, Belleza Castro 1995, Hardman de Bautista 1969: book 2, 8) call both the paternal and the maternal uncles *kăka* and *kaka*, respectively.

➢ *The origin of* kaka *in Quechuan and Aymaran*

Quechuan and Aymaran have at least 9 kinship terms in common, and this fact, along with many other vocabulary similarities, is likely to be interpreted as loans from Quechuan into Aymaran. But Hardman's claim (Pyle 1981: 89, editor's note) is that, as with *warmi* 'woman, wife,' *"in the majority of cases the direction of borrowing was from a Jaqi language to a Quechua language, particularly Cuzco Quechua."* And she continues: *"Because Quechua was the imperial language at the time of the conquest, any similarity anywhere has been regularly attributed to Quechua origins. This is not the case; imperial Quechua was a relatively recent imposition. The full complexity of the prehistory of language contact in the Andes is yet to be told."*

So it is not clear yet which language the word *kaka* was borrowed from, *if it was borrowed at all*. This term might well have been original in both language families.

2.4.2. Koko *in Mondé: ca. 2,000 BP*

According to Moore (2005: fig. 1), the Mondé language group, a subfamily of Tupían, is composed of 3 languages distributed between 2 branches. The first branch comprises Suruí, while the second one is divided between Salamãy and a third language including 4 dialects: Gavião, Zoró, Cinta Larga and Aruá. Anonby and Holbrook (2013), for their part, acknowledge the existence of 6 languages in Mondé: Salamãy, Aruá, Surui, Gavião, Zoró, Cinta Larga, but their classification only accounts for the 4 languages still alive: Gavião, Zoró, Cinta Larga and Suruí.[35] Their conclusion assumes that Zoró is more closely related to Suruí than it is to Gavião, and thus delineates two branches, with Gavião constituting the first one, and Cinta Larga and Suruí-Zoró constituting the second. Brunelli's opinion (1987: 157) is that the Surui separated from the other Mondé groups 1,500 years ago, while Macedo Brito (2005-2006) posits this split by ca. 2,000 BP. For their part, Anonby & Holbrook (2013: 29) assume, without much precision, that Gavião diverged *"very early"* from Proto-Mondé, whose homeland is posited along the Roosevelt and Aripuanã Rivers by Brunelli (1987: 157) or in the region of the Ji-Paraná River's headwaters by Campbell (1997: 199). The comparative material involving *KOKO* terms is compelling. Suruí, Gavião, Zoró, Cinta Larga *KOKO* terms are all address terms; we do not know if it is the same for Salamãy.

In the absence of knowledge of the sound correspondences existing in the whole family, but to take in account the recurring sounds and meanings exhibited in the following cognate set (table 10), we will just propose, an "inspectional reconstruction" preceded by the symbol $^{\#}$.

Table 10. Mondé comparative data arrayed according to Moore's internal classification

Proto-Mondé	$^{\#}$ ***kotkoot***	**MB, (EF)**
First branch		

---

[35] Their classification concerns the languages where field research was possible.

Suruí	*k*	*o*		*k*	*ó*			U (MB) ad.
Second branch								
Salamãy (Sanamaica)	*k*	*o*		*k*	*o*			U (MB)
Gavião of Rondônia (dialect of the third language)	*k*	*ó*	*t*	*k*	*ó*	*ò*	*t*	MB ad. (D. Moore, pers. comm.)
Zoró (dialect of the third language)	*k*	*u*	*t*	*k*	*u*		*t*	MB, FZ, GP ad.
Cinta Larga (dialect of the third language)	*k*	*o*		*k*	*ó*			MB, MBS, WF ad.

2.4.3. KOKO *in Nambikuáran: 2,807 BP (ASJP)*

The Nambikuáran languages. (Map: Wikipedia).

The Nambikuáran language family is composed of 2 major branches: Northern and Southern Nambikuáran, themselves including several languages (dialects), and one separate language, namely Sabanê (Eberhard 2009-1: 21). This family spans a territory covering the northwestern part of Mato Grosso and contiguous parts of the territory of Rondônia (Brazil). Directed by Professor Wetzel, a language study program was launched in 1998, notably involving the phonological and grammatical description of the languages from the 3 branches, and ultimately including a comparative study and a reconstruction of Proto-Nambikuáran (Telles & Wetzels n.d.: 236). Several doctoral dissertations since then have been completed, notably *Fonologia e Gramatica Latundê/Lakondê* (Telles 2002), *A Grammar of Sabanê* (Antunes de Araujo 2004), *Mamaindê Grammar: A Northern Nambikwara Language and Its Cultural Context* (Eberhard 2009). Another study concerning Sararé, a Southern Nambikuáran language, is in preparation (Borella).

The comparative list below displays the kinship terms referring to MB and EF collected from various languages of this family. Most forms display the root first two syllables, then a suffix *-nu* or *-ni* present in kinship terms (Eberhard 2009-1: 186), and the final nominal (referential) suffix *-su*, *-te*, *-tu* (Kroeker 2003: 12, 24, etc.; Eberhard 2009: 365-372, etc.). Not knowing what exact sound correspondences exist in the Nambikuáran family, we will again venture a tentative Proto-Nambikuáran form for MB/EF.

Nambikuáran comparative data

Proto-Nambikuáran[36] ***#-kṵ̃ŋkV-nVCV*** **MB, EF**; ***Southern Nambikuáran*** Halotesu *-kṵ̃ka-nusu* MB, EF; Kokozê (Juina Kitãulhú?) *kôkô-zu* MB; Waklitisu *-kuŋkú-nusu* MB; Elotasu *-kuŋkú-nusu* MB; ***Northern Nambikuáran*** Mamaindê *-kuhniru* MB; dialect b 1/2 (Lévi-Strauss 1948) *-kún-de* MF, EF; ***Sabanê*** Sabanê *kooka* MB, FF, EF.

2.4.4. KOKO *in Karirían*

Karirí is considered either as one single (extinct) language including several dialects, or as a language family composed of several languages. The dialects (languages) for which some documentation is available, and thus mentioned in our report are Kipeá, Dzubukuá, Sabujá and Pedra Branca (Kamurú). The presence of Karirían-speaking groups is attested ca. 1670 in the states of Paraiba, Pernambuco and along the rio São Francisco, at least upstream of Belém de São Francisco, in a place named Aracapá (Ouraccapa) (Martin de Nantes n.d. [published in 1706 or 1707]: 1, 2, 4, 22, etc.). In the preamble of his Portuguese-Karirí *Katecismo Indico da lingua Kariris*, the Capuchin missionary Bernardo de Nantes (1709: aiiij) gives two precious bits of information: first, that the language in which the catechism is translated was called Dzubukuá and spoken along the rio de São Francisco, thus confirming what Martin de Nantes previously reported. Second, in the same preamble, he also mentions that the language in which the Jesuit Mamiani wrote another *Catecismo* (1698), and also an *Arte de Gramatica*: (1699) was a different language named *Kippea*, and adds that the two *"nations"* were geographically separated by *"more than a hundred leagues"* (1709: aiiij). In 1818, Spix & Martius (1828: 615) found two Karirían-speaking communities living in the south of what is now the little town of Santa Terezinha in the state of Bahia (Costa Neto 2007: 60): the first one was established in *'Villa de Pedra Branca,'* (where the present-day village of Pedra Branca is situated), a toponym which became linguistically associated with the ethnic group, the second one was established "*a quarter of an hour*" south of this first settlement, in a place then named *'Caranquejo,'* and formed a group known as the Sabujás. Altogether these groups consisted of some 600 people. Spix & Martius (1828: 615) added that before they settled down in these two places, the Kariris were scattered in the woody mountains nearby – probably in the Serra da Jibóia.

---

[36] A subscript tilde ̰ marks a creaky voice (laryngealized) vowel.

Rodrigues (1986: 48-55, 1999: 164-206) considers that the Karirían language(s) belong(s) to the Macro-Jê trunk; Greenberg (1987: 84, 384), and Greenberg and Ruhlen (2007: 280) hold that Karirí is close to Tupian, including both of them in a putative Karirí-Tupi cluster, itself a component of the Equatorial stock. Kaufman (2007: 73) says that Karirí is an unclassified language.

The first mentions of *KOKO* kinship terms in Karirí (Kipeá) are found in Mamiani's *Catecismo* (1698: 197): *icucú* [ikuku] 'os tios,'[37] as well as in his *Arte de grammatica* (1699: 19): *cucù* 'tio.' In Bernardo de Nantes' *Katecismo* (1709: 91), displaying the Dzubukuá dialect (language), we find *i-cucu-te* 'os tios.' Von Martius, for his part (1867a: 359, 1867b: 215, 218) recorded *cuccu* and *cuccùh* 'Oheim, avunculus' (MB) for Sabujá, and *cuccuh* 'avunculus' (MB) for the Kariris living in Pedra Branca.[38] The term *kuk(k)u(h)* MB is consequently not a newly acquired or borrowed term in this language family, and can be postulated in Proto-Karirían.

2.5. **KOKO** *in other South American families and languages isolates*

➢ *Families*

Barbacoan: Guambiano *kásuko* U; Colorado *ʔaʰ'ko* B; **Chocoan:** Emberá Catio *káu* D; Choco (Emberá) *kau* D; Epena *'kʰau* D, girl; Noanamá *ka* D, *kúi* Z, *kawá* EM; **Cholonan:** Xibito *kotk* F, *keek* M; **Chonan:** Puelche *ukɨči* B; Tehuelche *koka* F, dueño, *kok'an* madrina, *goo* B, Z, *qon* GM; Selknam *ho'o* GF, *hoho'nh* GM; **Jêan:** ***Northeastern Jê:*** Panará (Kren Akarore) *kokri-pia* EM; ***Southern Jê: *kãke* eB, **kakrã* MB, EF**; Kaingang (Toldo das Lontras, Nonoai, etc.) *kakre* MF, MB, WF; Kaingang (Rio Grande do Sul) *kakrö* MF, MB, EF, *kɛ̃ke* eB, eZ; Kaingang (Palmas, Parana) *kakrá* ~ *kakrẽ̱* EF, *kãke* eB, eZ; Kaingang (Serra do Chagu, Parana) *ikãke* B; Kaingang (Duque de Caxias) = Xokléng *kɔkla* ceremonial F = MB, *kaké* relative, cousin, but not B, rarely A; Ingáin *kau* U; **Jivaroan:** Jivaro Achuar *kai* (♀)B, (♀)Z; Huambisa *kair* (♀)Z; Aguaruna *kaig* (♀)Z; **Kamakanan:** Kamakán *gkoong* GF, MB; **Karajá language area:** Karajá *hi* (♀)eB, *ixi* yB;

---

[37] The prefix *i-* is the 3rd person possessive marker.

[38] Adam (1897) gathered a lexical comparative material based on the reports done about the Kipeá and Dzubukuá dialects, by Mamiani and Bernardo de Nantes respectively, and about Sabujá and Pedra Branca dialects, done by von Martius. Adam, who faithfully reported von Martius' transcriptions, remarked that the final vowels in Pedra Branca and Sabuyá were frequently affected by an aspirated *h* (1897: 6). His comparative glossary also shows the occurrence, notably in the transcriptions of the second syllable in Sabujá and Pedra Branca languages (henceforth S. and P.B.), of geminated consonants *c* [k], *m*, *t*, *b* or *p*, not present in Kipeá and Dzubukuá (henceforth K., D.): K., D. *cucu* MB, S., P.B *cuccuh* MB; K., D. *ami* 'food', P.B. *ammih* id.; D. *bati* star, P.B. *battüh* id.

Krenákan: Botocudo *ñ-gikę̊n* (♀)F, (♀)EF; **Maxakalían:** Maxakalí *xuxyã* MB, EF, FF, MF, FFF, MFF, *xukux*, FZ, EM, GM, MMM; **Lule-Vilelan:** Lule-Tonocote *kue* FZ; **Makúan:** Hupda *ʔuw* GF, *ʔɔh* GM, *ʔoʔ* eB; Nadëb *óów* GF, *hooñ* GM; Puinave *hika* A, *-u'ɨi* EF; **Mascoyan:** Sanapaná Enlhet dial. *haawok eŋak* eB; Mascoy (western) *koko* MB, *hawok* eB, BS; Mascoy (Lengua) *hawok* eB, BS; **Matacoan:** Chulupí (Nivaclé) *k'ucaʔx* old man, *k'ucxa* old woman, *yi-kakt'ech* WF, *yi-kakt'e* my WM; Chorote *kihil* ~ *kíhi* FF, old man; Macá *keiuk* B, *k'ucax* old man, *-kewket* EF, *-kewket-iʔ* EM; Mataco (Wichí Lhamtès Vejoz) *-kó* M, *kawaklá* ZH; **Muran:** Pirahã *kaai* D; **Paezan:** Paez *kắʰka* MB, MZH; **Qawasqaran:** Alakaluf (Qawasqar) *xoyko-las* old woman; **Sáliban:** Sáliba *koko* man, *tihoho* A; **Timotean:** Timote *kuxioy* GM; Mucuchí *kaak* man; **Tukanoan:** Coreguaje *cu'eʉ* GF, *cu'eo* GM; Siona *kʷi* MB, MBS; Secoya *kʷi* MB, MBS; Orejón (Mai Huna) *gũ* MB; Barasana *gagü* eB, *gago* eZ; **Tupían:** ***Western Tupían: Arikém*** Arikém *u-këra* (♂)B; Karitiána *syky* (♀)B; ***Mondéan*** **#*ko(t)koo(t)*** MB, (EF) (see 2.4.2); ***Puruborá*** Puruborá *kɔkɔ* MB; ***Ramaráman*** Káro (Urukú) *iká* eZ, *hakóiga* H; ***Tuparían*** Tuparí *-ike* (♂)eB; ***Eastern Tupían: Mundurukún*** Mundurukú *ukipid* (♂)yB, *o-köt-köt* yMZS, yFBS; ***Mawé*** Mawé (Maué) *uy-kiwid* (♀)B; ***Awetí*** Awetí *i-kywýt* (♀)B; ***Tupí-Guaranían*** **#*kiwyra* (♀)B;** *Guaranían branch*, Guayakí (Aché) *kyvangi* (♀)B; Xetá (Shetá) *koti* B; Chiriguano *kiuỹ́*, *quigüi ~cekigüi* (♀)B; *Guaráyoan branch* Sirionó *áke* SS, DD; *Tupí branch* Tupinambá *kybyra* (♀)B; Cocama-Cocamilla *kiwirą* (♀)B; *Teneteháran branch* Tenetehára (Guajajára) *he-kiwyra* (♀)B; Parakanā *-kywyra* B; Tapirapé *che-kyvuyra* (♀)B; *Xingu branch* Araweté *čiwi* (♀)B; *Kawahíb branch* Tenharín *kuvyr* (♀)B; Kamayurá *ie-kywýt* (♀)B; Kayabí *kɨwɨt* (♀)B; Apiaká *erarkuireè* (♀)B; *Northern Tupí-Guaranían branch* Ka'apór *ihēkywyr* (♀)B; Wayampí *kakay* eB; **Witotoan:** Nonuya *hōkhá* man, *hohēkŏ* WB, HB *hokohokŏ* WZ, HZ; Ocaina *hoh́ə̆* man, *hăhāhō(h)ə̆* WB, HB, *kōhōhŏ* cousin; Witoto *mota uaikika* GF = F + old man; **Zamucoan:** Ayoreo *axai* (♀)B, *axu* MZ; Chamacoco *-okok* EF; Ebidoso *-ohot* EF; Tumereha *-ohó* 'my EF'.

➢ ***Language isolates***

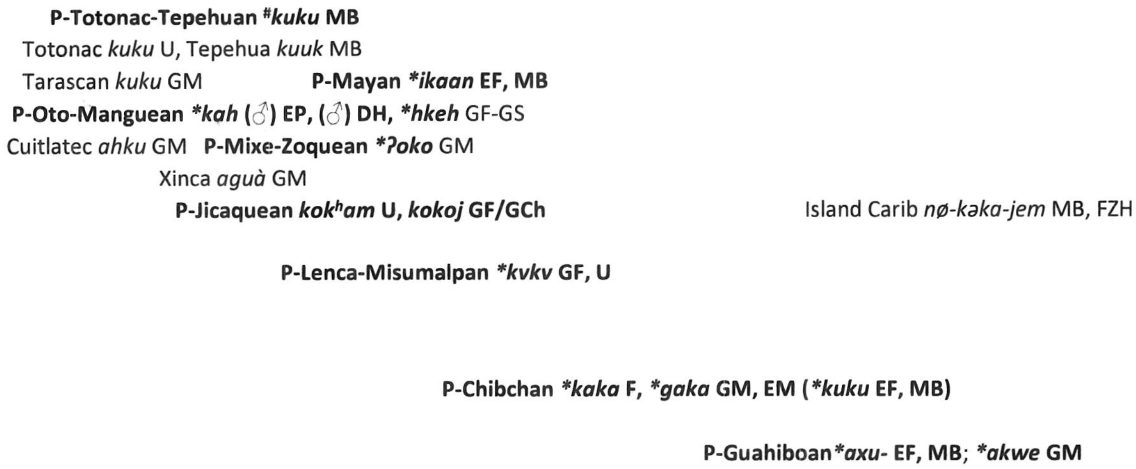
Aikaná (Masaca) *kokomai* U; Kaliana (Sapé) *ma-kohai* ~ *ma-ku̲hai* my GF, *kohai* old man; Camsá *-ki* MB; Canichana *eu-axa* my GF; Cayuvava *-kiče* U; Cofán *tõʔto* U; Kapixaná (Kanoe) *kṹkṹì* man, *keke* GM, *ūkū* HBD, WBD; Koayá (Kwaza) *ha'kai* GP; Mapudungu (Mapuche) *hueku* MB, *kacü* FZH, *kuku* FM, FFZ; Mochica (Chimú) *kokəʊd* (♂)eZ, (♂)A, *ikiş̧* EM; Movima *aiku* A, *ákai* eB, eZ; Munichi *tcʔaʔa* GF; Otí *koaka* B; Taushiro *'ukku* (♂)Z; Trumai *koko* MZ, FBW, *aoké* EM, FZ; Urarina (Simacu) *ka-kaun* FZ; Warao *ku*, *da -ku* MB; Yaté (Fulnio) *ī-xi* B.

3. The kinship term *KOKO* in Meso-America[39]

Among the first records of *KOKO* kinship terms in Meso-America, we find Tarascan *cucu* [kuku] GM, by Gilberti (1901 [1559]: 33), Mixe *oc(taac)* [ok(taak)] GM, A, by Quintana (1733: 80-81), Totonac *coco* [koko] U, by Zambrano Bonilla (1752: distintos p.1, *nombre de parentesco* p. 3).

---

[39] We use the term Meso-America to stay in line with Kaufman's (2007b) terminology.


P-Totonac-Tepehuan *#kuku* MB
Totonac *kuku* U, Tepehua *kuuk* MB
Tarascan *kuku* GM
P-Mayan **ikaan* EF, MB
P-Oto-Manguean **kah* (♂) EP, (♂) DH, **hkeh* GF-GS
Cuitlatec *ahku* GM
P-Mixe-Zoquean **ʔoko* GM
Xinca *aguà* GM
P-Jicaquean *kokʰam* U, *kokoj* GF/GCh
Island Carib *nø-kəka-jem* MB, FZH
P-Lenca-Misumalpan **kvkv* GF, U
P-Chibchan **kaka* F, **gaka* GM, EM (kuku* EF, MB)**
P-Guahiboanaxu-* EF, MB; **akwe* GM**

Map 2. Approximate locations across Meso-America of *koko* proto-forms (in bold), with additional terms from individual languages showing the extension of the distribution of *koko* words.

3.1. The geographic and linguistic distribution of KOKO *in Meso-America*

We have studied 178 languages/terminologies[40], living or extinct, from this region. *KUKU ~ KAKA* forms have been found in all of the 11 genetic units acknowledged in Kaufman's (2007) classification (map 2, table 11). The Lencan and Misumalpan languages mentioned in this classification have been already presented in subsection 2.3.7.

3.2. The reconstruction of KOKO *in Meso-American linguistic families*

From now on, the linguistic material will be presented by means of tables. These tables display the languages, the language families with their supposed time-depth, the 'stocks' when there is some consensus about their existence, and the reconstructed forms when available. These reconstructed forms haven't been evaluated like their South American counterparts, because of the space that such a discussion would entail. We will also venture a few "inspectional reconstructions" when the comparative tables display very similar forms.

---

[40] Lenca-Misumalpan languages not included.

Table 11. Meso-American comparative data[41]

	Stocks, families, languages	Masculine denotata	Feminine denotata	References
1	***Tequislatec-Jicaque (Tol)***	***Proto-Jicaque *kokʰam* U, **(ko)koy* (♂)GF**	***Proto-Jicaque *(ku)kus* D**	**R: Campbell & Oltrogge 1980**
	Tequistlatec	*kayʔáyiʔ* F, *-kúweʔ* HF		Turner & Olmsted 1966
	Jicaque	*kokoy* FF, (♂)GC, *kik* (♀)GC, *kokham* U	*kukus* D, SW, *kik* (♀)GC	Neuenswander 1977-81
	Western Jicaque	*cui*, *coho* GF, *goain* U	*gut* D	Campbell & Olt. 1980
	Montaña de la Flor	*kokoy* GF, *kokʰam* U, *ŋgokʰam* my U	*kukus* ~ *kuku* D	Campbell & Olt. 1980
	Cabeza de Vaca	*kokoy* old man, *kokam* U	*kukus* D	Campbell & Olt. 1980
	Lagunita	*cocoy muy* old man, *uncacom* U		Campbell & Olt. 1980
	Lean and Mulia	*cocoy* old man	*kukus* D	Campbell & Olt. 1980
	Jicaque de Yoro	*cocam* U	*cucusuay* D, *gucús* DinL	Campbell & Olt. 1980
2	***Totonac-Tepehuan: 2,500 BP?***	***#kuku* U (MB)**		**University of Alberta LDRC, R: AM&PB**
	Totonac (18th cent.)	*coco* [koko] U, *gag chin* (♀)eB		Zambrano Bonilla 1752
	Filomeno Mata Totonac	*kʰúkʰu̥* U, *-kukustá'ḁ* BinL		McFarland 2009
	Totonac of Xicotepec de Juárez	*ix-kuk* thy U		Reid & Bishop 1974
	Upper Necaxa Totonac	*ku̯k* one's U, pl. *ku̯kúnu'*		Beck 2011
	Totonac Papantla	*cucu* [kuku] U		Aschmann 1973
	Totonac de la Sierra	*cucu* [kuku] U		Aschmann n.d. (2008)
	Coyutla Totonac	*kuku* U		Aschmann n.d. (2000)
	Misantla Totonac	*kin-kuk* my U		MacKay 1999
	Huehuetla Tepehua	*kuuk* MB		Smythe Kung 2007
	Tlachichilco Tepehua	*kin-kuku̥* my U		Watters 2010
	Tepehua de Pisaflores	*kin-kuku* [kíŋ kúku̥] my U		MacKay & Trechsel 2013
3	***Mixe-Zoquean: 3,000 BP***		****ʔoko* GM, (GC)**	**Campbell & Kaufman 1976, R: Wichmann 1999**

---

[41] For reason of space, the references – except a few in the general references – have been only reported by author names and publication dates. Complete references can be obtained from the first author: a.matthey@free.fr, or found online at language-kinship.org (click on tab Databases). References of reconstructed terms (in bold) marked * or # can be found in the general references.

Stocks, families, languages	**Masculine denotata**	**Feminine denotata**	**References**
Mixe (Totontepec)		*ok* GM, GD	Beals 1945
Mixe (Coatlan)	*ʔahč* (♂) older male relative, *ʔoknï°hš-* GS	*ʔoknï°hš-* GD, *koʔu* stepM	Hoogshagen & Merrifield 1961
Mixe (Metaltepec)	*ok* GS	*ok* GD	Beals 1945
Mixe (Juquila)	*okunk* (♀)GS	*ok*, *oktaak* GM, *oknox* (♀)GD, *xoicx* (♀)EM	Radin 1931 (Quintana 1733), Beals 1945
Zoque	*oko* MF, MGP, *oko unk* GS	*oko han* GD	La Grasserie 1898, Radin 1931
4 ***Mayan: ca. 4,200 BP***	***ikaan* (♂)MB, (♂)FZH, (♂)MGF, WF, **ikaaq* cousin**		**Kaufman 1976, R: Wichmann & Brown n.d., Kaufman & Justeson 2003**
Huastecan			
Huastec	*koko* tutor		Radin 1931, Schuller 1928
Huastec Western (Tancanhuitz)	*itxaan* U, *itxak'* nephew	*txanuub* A	Kaufman & Just. 2003
Yucatecan-Core Mayan			
Yucatecan			
Epigraphic Mayan	*ʔichaan* MB		Kaufman & Just. 2003
Maya (16th – 18th centuries)	*akan* (♂)MB, (♂)step-MB (♂)FZH, (♂)MGF	*n-okhó* WM, WEM, *kik* eZ, (♂)MBD (if older), (♂)SWM	Beltran 1742, Radin 1925, Eggan 1934
Yukateko (Mopan)	*acan* MB, *(a)chak* male kin		Kaufman & Just. 2003
Lacandon	*akän* MB, ZS, WF	*kik* eZ, *chiich* MM	Boremanse 1979
Core Mayan			
Ch’oltiʔ	*ichan* U, *ichoc* nephew		Kaufman & Just. 2003
Ch'ol	*ichan* MB	*ichak'* FZW	Kaufman & Just. 2003
Choti	*ičan* MB		Breton 1919 (1695)
Tzeltal	*ič’an* MB, MBS, MBSS, etc., *ichan* MB, *ichac* male kin	*čuču’* FM, MM, MBW	Sousberghe *et al.* 1962, Kaufman & Just. 2003
Tzotzil	*ʔičin* MyB, *ichok'* male kin	*čič* MBW, *čičil* GM	Schuller 1924, Hopkins 1969
Q'anjob'alan-Chujean			
Q'anjob'al	*ikan* MB, *ikan-ej* U	*ikan-ej* A	Kaufman & Just. 2003
K’ichean-Mamean			
K’icheʔ	*ican* MB		Kaufman & Just. 2003
Quiche	*v-igan* my MB, *ikaan* U, *r-ikaaq'* nephew	*r-ikaaq'* niece	Breton 1919 (1787), Kaufman & Just. 2003
Chichicastenango (Quiche dial.)	*r-ikaan* U, *r-ikaaq'* male cousin, (♀)nephew	*r-ikaaq'* (♂)niece	Kaufman & Just. 2003
Chicaj	*r-ikaan* U, *r-ikaag'* nephew	*r-ikaag'* niece	Kaufman & Just. 2003
Sipakapense	*kan* U, *ikaq’* nephew	*ikaq’* niece	Kaufman & Just. 2003
Sakapulteko	*kyaan* U		Kaufman & Just. 2003
Tz'utujiil	*ityan* U		Kaufman & Just. 2003

	Stocks, families, languages	**Masculine denotata**	**Feminine denotata**	**References**
	Kaqchikel	*v-ikan* my MB, *ikan* MB	*v-ikan* my MZ	Breton 1919, Kaufman & Just. 2003
	Palin Pocomam	*ikaan* U, *ikaq'* nephew	*ikaan* A, *ikaq'* niece	Kaufman & Just. 2003
	San Luis Jilotepeque	*ikaan* U	*ikaan* A	Kaufman & Just. 2003
	Poqomchiiʔ	*ikan* FB, MB, *ica7k* male kin	*ikaan* A	Kaufman & Just. 2003
	Uspanteko (Musre)	*r-ikaan* U		Kaufman & Just. 2003
	Q'eqchiʔ (Kekchí)	*ikan* U, *ikak'bej* nephew	*Ikan na7* (*ikan na'*) A, *ikak'bej* niece	Sedat 1955, Kaufman & Just. 2003
	Western Q'eqchiʔ (Coban, Chamelco)	*ikan* MB, *ik'aq* nephew	*ik'aq* niece	Kaufman & Just. 2003
	Mam /qyool/	*ikiam* U	*ikiak* niece	Kaufman & Just. 2003
	San Idlefonso Ixtahuacan (Mam dial.)	*t-kyaan* (♂)male cousin	*t-kyaan* (♂)female cousin	Kaufman & Just. 2003
	Teko (Cuilqueno)	*n-chaan* U		Kaufman & Just. 2003
5	**Otomanguean:** ***ca. 4,000 BP?, 6,400 BP?, 7,400 BP?***	***hkeh* GF, **kha* male affinal kinman, **kuHn* Sib**	***kwaHn* (♀)eZ, **kuHn* Sib**	**Kaufman & Just. 2009, Campbell 1997, ASJP, R: Merrifield 1981**
6	***Lenca-Misumalpan: 7,200 BP* (see section 2.3.7)**	***kVkV* GF, (MB), (EF)**		**Constenla Umaña 2002, R: id.**
	Isolates			
7	Tarascan		*cucu* [kuku] GM	Radin 1925 (Gilberti 1559)
8	Xinca		*aguà* GM	Breton 1919 (1770)
8	Cuitlatec	*ahcí* GF	*ahkú* GM	Drucker, Escalante, Weitlaner 1969
9	Huave	*-kóh* eB, *òkwáàk* WF	*-kóh* eZ, *òkwáàk* WM	Diebold Jr 1966

4. The kinship term *KOKO* in North America

Among the first *KOKO* kinship terms recorded during the postcontact period are those from the Algonquian languages: Wampanoag (Natick) *kokummes* thy A, *kokummes* thy GM (Trumbull 1903, from Eliot's Bible, 1663), Montagnais *n8k8mis* [nookoomis] FB (Silvy ca. 1678), Abenaki *n8'k8mes*, *n8'k8mi* [nookoomes, nookoomi] my GM, my A, *n8'k8m* [nookoom] my UW, my UZ (Rasles 1691 [1833]: 498-9), Miami *n8c8ma* [nookooma] GM (Gravier 1700). Let us also mention *caca* [kaka] FM in Cohahuiltecan by Garcia (1760), quoted in Romney (1967: 229).

4.1. The geographic and linguistic distribution of KOKO *in North America*

252 languages (or dialects) out of the 360 North American languages and dialects of our sample display *KOKO* kinship terms. As shown on map 3, *KOKO* terms pervade the entire North American territory and are found in 39 out of the 58 genetic units considered in Golla *et al.* (2007).

Point Barrow *akaakuk* FB East Greenland *akka* FB
Seward Pen. *aaka* M N. alaskan Nunamiut *aaka* M West Greenland *akkaq* FB
Kangianermiut *akkakkan* FB, *akkan* M Pond Inlet *akak* FB
P-Athapaskan **ank̓ayʔ* MZ Simpson Peninsula *akkaq* FB Iglulik *aqak* FB
Eyak *aqak* MeB, *aqakcia* MyB
Tlingit *kák* MB, *akhkē* B Southampton Isl. *akak* FB
Baker Lake *akak* FB Sugluk *atkak* FB
Upper Kazan River *akkaq* FB
Eskimo Point *akak* FB Port Harrison *akkak* FB
Haida *qā* MB, *(ti)ka(-gha)*
Great Whale Riv. *aka'* FB
P-Salishan **sxa'xa* EF, DH
Chinook *gaga* GF
P-Northern Sahaptian **ka'ka* MB, *na-kakas* 'my MB'
Coos *axaax* MB **P-Algonquian **n-ookho* GM**
Yurok *ne-kwa* EF, EM, Karok *xokam* U
Coast Yuki *kaha* MB/**P-Wintuan *#kije* MB, FZH/P-Maiduan *#kaka* MF, MB, FZH**
P-Miwokan **kaka* MF, MB, MBS, *#kawu* FZH, ZH
Pomo *kakam* MM **P-Siouan **khų́* GM, EM**
P-Yokutsan **ʔakas* MB
Zuni *kaka* MB, *kuku* FZ Kiowa *q`u* (♂) GF, (♂)GS Catawba *koko* U
Isleta *k'iwʔu* FZ
P-Caddoan **kaʔ* GM
Natchez *gaga* eB
P-Uto-Aztecan **kwaʔa* MF, **kaʔ* FP, **ko* eZ Chitimacha *koʔ* A
Cohahuiltec *kaka* FM, *ku-anax* MeB
Classical Nahuatl *kohkoli* GF
Nahuatl Xalitla *kohkol-* GF Nahuatl Mecayapan *-koko* U, eB
Pipil *kuhkul* old man (GF)

Map 3. Approximate locations across North America of *koko* proto-forms (in bold), with additional terms from individual languages showing the extension of the distribution of *koko* words.

4.2. KOKO *in languages and language families of North America*

As was done for Meso-America, the linguistic material is presented in a table.

Table 12. North American comparative data[42]

NB: Languages are arrayed from north to south

	Stocks, families	**Masculine denotata**	**Feminine denotata**	**References**
1	***Eskimo-Aleut***			
	Eskimo		***aka* eZ, M**	**R: EHL**
	Inuit	***akkak* FB**	***a(a)kkaa* M, eZ**	**R: Fortescue *et al.* 1994, EHL**
	Inuktitut			
	Iglulik	*aqak* FB		Stevenson 1964
	Simpson Peninsula	*akʌ˚q* FB		Birket-Smith 1928
	Melville Peninsula	*a˚kaη* FB		Birket-Smith 1928
	Pond Inlet	*akak* FB		Stevenson 1964
	Chesterfield Inlet	*akak* FB		Stevenson 1964
	Upper Kazan River	*akʌ˚q* FB		Birket-Smith 1928
	Southampton Island	*akak* FB		Stevenson 1964
	Rankin Inlet	*acug* FB		Stevenson 1964
	Baker Lake	*akak* FB		Stevenson 1964
	Eskimo Point	*akak* FB		Stevenson 1964
	Cumberland Inlet	*ŭk'-ŭgŭh* FB		Dall 1877
	Frobisher Bay	*akakulu* FB		Stevenson 1964
	Lake Harbour	*akakulu* FB		Stevenson 1964
	Sugluk	*atkak* FB		Stevenson 1964
	Port Harrison	*akkak* FB		Stevenson 1964
	Labrador	*akka* FB		Rasmussen 1941
	Great Whale River	*aka'* FB		Stevenson 1964
	West Greenland	*akkak* FB		Gessain *et al.* 1982
	Thule (North Green.)	*ak'äk* FB		Birket-Smith 1928
	Northumberland	*uk'-kä* FB		Morgan 1871
	East Greenland	*akka* FB		Gessain *et al.* 1982
	Inupiatun			
	North Alaskan Inupik		*āka* M, *āxxā* eZ	EHL
	Point Barrow	*akaakuk* FB		Stevenson 1964
	Point Hope etc.	*aqaga* ~ *akaakaŋa* FB, MB	*aakaŋa* M	Heinrich 1960
	Nunamiut	*akanakan* male cousin	*aaka* M, *akaqarak* stepM	Pospisil & Laughlin 1963, Pospisil 1964
	Kangianermiut dialect	*ak˚ak˚an* FB	*a˚kan* M	Rasmussen 1941
	Bering Straits	*akaakaŋa* MB		Heinrich 1960
	Seward Peninsula		*āka* M	EHL
	Yupik		***áka* eZ**	**R: EHL**
	Alutiik Yupik		*ākāq* eZ	EHL
	Norton Sound		*ākaq* M	EHL
	Sirenik		*ākaχ* eZ	EHL
	Chaplino		*áká* eZ	EHL
	Naukan		*ákaq* eZ	EHL
	Aleut			
	Eastern Aleutian and Alaskan Islands		*kúkaq* GM	Geoghegan 1834

---

[42] See also note 25.

	Stocks, families	**Masculine denotata**	**Feminine denotata**	**References**
2	***Nuclear Na-Dené: 8,532 BP***			**ASJP**
	Tlingit	*-k'a'k'* MB, *k'ik'* (♂)yB, *'ik'* (♀)B		Mayer-Durlach 1929
	Eyak-Athapaskan: 3,500 BP, 4,234 BP			**Campbell 1997, ASJP**
	Eyak			
	Eyak	*aqaq* MeB, *aqaq-cia* MyB	*aki*^{n} (♀)FeZ, *aki*^{n}*-cia* (♀)FyZ	Birket-Smith & De Laguna 1938
	Athapaskan: 2,500 BP		***-anḱayʔ* MZ**	**Krauss 1973, R: Hoijer 1956**
	Tanaina		*-o°ka* MZ	Hoijer 1956
	Anvik		*-q́ói* MZ	Hoijer 1956
	Tanana		*-age* MZ	Hoijer 1956
	Kutchin		*-ɛ̀°ḱái*, *aki* MZ	Hoijer 1956
	Yellow Knife		*-akrea* MZ, stepM	Hoijer 1956
	Dene (Chipewyan)		*-ą̀ḱíyɛ̀* MZ	Hoijer 1956
	Carrier		*-aḱ °ai* MZ, ZD, *akwal* my M	Goldman 1941
	Tolowa		*kaka* M, *onkai* MZ	Gifford 1922, Hoijer 1956
	Hupa		*-ŋḱaiʔ* MZ	Hoijer 1956
	Wailaki		*-iŋkait* MZ, FBW, stepM	Hoijer 1956
	Lassik		*unkai* MZ, FBW, stepM	Hoijer 1956
	Sinkyone		*unkai* MZ, FBW, stepM	Hoijer 1956
	Navajo		*-íḱàʔí* MZ, stepM	Hoijer 1956
	San Carlos		*-íḱą̀°ʔé* MZ, (♀)SC, stepM	Hoijer 1956
	Chiricahua		*-íḱą̀°* stepM	Hoijer 1956
	Mescalero		*-íką̀°è* MZ	Hoijer 1956
	Jicarilla		*-íą́ḱ°* FBW, stepM	Hoijer 1956
	Lipan		*-íḱá°* stepM	Hoijer 1956
	Kiowa Apache		*-k'àʔá* MZ	Hoijer 1956
	Apache		*-k'aʔaʕ* MZ	Donald & Tighe 1987
3	***Algic: 7,200 BP?, 4,000 BP?, 5,554 BP?***		$^\#$***-oko-* GM**	**Swadesh, Golla 2007, ASJP, R: AM&PB**
	Wiyot		*yi-d-oko-tck* GM, EGM, *hakwi* E (endearing)	Gifford 1922
	Yurok	*ne-kwa* EP	*ne-ku-ts* GM, *kok* ad. M	Gifford 1922
	Algonquian: 3,000 BP, 3,343 BP		***-okko-*, *n-oohko-* ad. GM, (EM)**	**Proulx 1981, ASJP, R: Sapir 1922, Wheeler 1982**
	Central Algonquian			
	Montagnais	*nookoomis* my FB	*noqu'm* my GM, *noqumu'c* my FZ, *nəqa'wi˚* my M	Silvy ca. 1678, Speck 1918
	Cree	*no˚hkomis* my FB	*no˚hkom* my GM, *no˚hkoh* ad.GM, *nika˚wiy* my M	Hockett 1964
	Prairie Cree	*no'komis* my stepF	*noh'kome'* my GM, my EGM, *n'gäwe* my M	Morgan 1871

Stocks, families	**Masculine denotata**	**Feminine denotata**	**References**
Plains Cree	*n'o'komis* my MB	*nokûm* my GM, *n'qawi* my M	Skinner 1914
Woods Cree	*no'komish* my stepF	*nokome'* my GM, my EGM, *n'gäwe* my M	Morgan 1871
Lowlands Cree	*no'komis* my stepF	*nokome'* my GM, my EGM, *n'gäwa* my M	Morgan 1871
Ojibwa		*nokomiss* my GM, *no°kko°* ad. GM, *ninga* my M	Trautmann & Barnes 1998, Hockett 1964
Ottawa		*nokomis'* my GM, my WGM, *n'gus'-sheh* my M	Morgan 1871
Miami		*nōkomä'* my GM, my WGM, (♂)EM, *nin̤geah* M	Gravier 1700, Morgan 1871, Costa 1999
Peoria		*nokomä'* my GM, my WGM, *nin̤geah* M	Morgan 1871
Piankeshaw		*nokomä'* my GM, my WGM, *ningeah* M	Morgan 1871
Kaskaskia		*nokomä'* my GM, my WGM, *negeah* M	Morgan 1871
Weaw		*nokomä'* my GM, my WGM, *negeah* my M	Morgan 1871
Kickapoo		*nō'komeza* my GM, *nō'ko* ad. GM, *nō'komḁ* my EM, *negya* my M	Dyneley Prince 1913
Potawatomi		*nok°məs°* my GM, *nkye* my M	Hockett 1964
Menominee		*no°hkomɛh* my GM, *no°hkoq* ad. GM, *nekiah* my M	Hockett 1964
Fox		*no°'gome's*^{a} GM, *ano°hko* ad. GM, *no°hkoma* my EM, *nogum*^{a} my EM, *negy*^{a} my M	Eggan 1937, Hockett 1964
Sauk		*no'komis* my GM, my WGM, *noko'mä'* my EM, *năkea'* my M	Morgan 1871
Shawnee		*nokomqθa* my GM (form uncertain), *nokomeea* my GM, my WGM, *nakeah* my M	Morgan 1871, Hockett 1964
Eastern Algonquian			
Micmac		*no'ɣomi°tc* my GM, *no'ɣəmi°tc* my stepM, *nki°tc* my M	Speck 1918
Malecite		*no''kəməs* my GM, *no'°kəməs* my stepM, *no''kəm* my FZ, my MBW, *ni°'gəwus* my M	Speck 1918
Passamaquoddy		*no''kəməs* my GM, *no'°kəməs* my stepM, *no''kəm* my FZ, my MBW, *ni°'gəwus* my M	Speck 1918

Stocks, families	**Masculine denotata**	**Feminine denotata**	**References**
Western Abenaki (St Francis)		*nōkōmes* my GM, my A, *nō'kōm* ad. FZ, *nokəmə's* my GM, my FBW, my MZ, *nok'əmə's* my stepM, *nokə'm* ~ *nuku'm* my FZ, my MBW, *nigawəs* my M	Rasles 1691, Speck 1918
Eastern Abenaki (Penobscot)		*no''kəməs* my GM, my stepM, *no*$^{o'}$*'kəm* my FZ, my MBW, *ni°ga°'wəs* my M	Speck 1918
Wampanoag (Natick)		*kokummus* thy GM, *kokummes* thy A, *okummes* A, FBW *-ookumiss-* GM	Trumbull 1903, (from John Eliot's Bible)
Mohegan (and Pequot)		*nohkumihs* my A	Dictionary (council of elders): 2006
Mahican (Morgan: Mohegan)		*noome'* my GM, *n'guk* my M	Morgan 1871
Munsee		*nano'home* my GM, *nainguk* my M	Morgan 1871
Delaware		*nooho°me'* my GM, *n'gähase* my M	Morgan 1871
4 ***Wakashan: 2,781 BP***			**ASJP**
Kwak'wala (Kwakiutl)	*gágas* your GF	*gágas* your GM	Grubb 1977
Hesquiat (Central Nootka)	*k*^{w}*a°ʔoc* GGP(S), *ʔu°ʔa°c̀u* middle B	*k*^{w}*a°ʔock* GGP(D)	Fleisher 1984
5 ***Salishan: 3,827 BP***	***qá-* eB, $^{\#}$*s-xa'xa* EF, DH, SW**	***kíx* eZ**	**ASJP, R: Kinkade 1992, Morgan 1980**
Coast Salish			
Squamish	*skāk°* yB, *kō'kpi* GF	*skāk°* yZ	Boas 1890
Bilqula	*kō'kpi* GF		Boas 1890
Twana (Skokomish)	*sx̥ax̥áʔ''* EF, DH		Elmendorf 1946, Morgan 1980
Snohomish	*ska* eB		Haeberlin & Gunther 1942
Puget Sound Salish (Nisqualli)	*sxa'xa* EF, DH, *qαq*^{hu} ad. eB	*sxa'xa* SW, *qαq*^{hu} ad. eZ	Ballard 1935
Upper Chehalis	*k'*^{w}*ú-* F	*k'*^{w}*ú-* M	Kinkade 1992
Lower Chehalis		*kə́h* M	Kinkade 1992
Cowlitz	*x̱*^{w}*aʔɬ* eB	*x̱*^{w}*aʔɬ* eZ, *káyʔ* GM	Kinkade 1992
Quinault		*kə́* M	Kinkade 1992
Interior Salish			
Shuswap	*sqā'qoa* EF, *k°a'tsk°a* eB	*ka'ku* eZ	Boas 1890
Spokane	*s-x̥aʔ-x̥éʔ* EF		Morgan 1980
Okanagan-Colville	*sxăxăʔ* ref. & ad. EF, *k'ík'waʔ* ref. & ad. MF, MPB, *ɬqáqcaʔ* ref. eB, *qácaʔ* ad. eB, *xaʔxʔít* elder, ancestor	*kíkxaʔ* ad. eZ	Mattina & Jack 1992

	Stocks, families	**Masculine denotata**	**Feminine denotata**	**References**
	Flathead Salish (Kalispell)	*sxa̯’x̯é’* EF, *qáqce’* (♀)eB, *qécč* (♂)eB	*qáqe’* ~ *qáx̥e’* MZ, *sḱúḱʷi* FZ	Krueger 1961
6	***Siouan-Catawban: 4,000 BP?, 6,523 BP?***			**Campbell 1997, ASJP**
	Siouan: 3,000 BP, 3,169 BP		***khų́* GM, EM**	**Campbell 1997, R: Matthews 1959**
	Hidatsa	*áka* (♂)eB, (♂)MB	*kú* GM, HM etc.	Matthews 1959
	Crow	*īke* (♂)eB, (♂)MB	*wa-kú-kāte* ad. older female relative	Matthews 1959
	Mandan	*ų́ka* (♂)eB		Matthews 1959
	Assiniboine		*kų́* EM, *kų́-ši* GM etc.	Matthews 1959
	Santee		*khų́* EM, *khų́-ši* GM, etc.	Matthews 1959
	Teton		*khų́* EM, *khų́-ši* GM etc.	Matthews 1959
	Winnebago		*kųų-nį́k* ad. GM, EM, etc.	Matthews 1959
	Iowa		*kó-ni* GM, EM etc.	Matthews 1959
	Oto		*kó-ni* GM, EM etc.	Matthews 1959
	Omaha		*ką́* GM, EM, etc.	Matthews 1959
	Kansa		*kǫ́* GM, EM, *kú* id. (respectful)	Matthews 1959
	Osage		*kǫ́* GM, WM, *kó* id. (respectful)	Matthews 1959
	Kwapa		*ką́* GM	Matthews 1959
	Biloxi		*kąką́* GM, EM	Matthews 1959
	Ofo		*kó-ni* GM	Matthews 1959
	Tutelo	*kukä’k`* ~ *kokowa°’nα* GF	*kų́* GM	Matthews 1959, Speck & Schaeffer 1942
	Catawban			
	Catawba	*kokó* U	*tcutcú* GM	Speck & Schaeffer 1942
7	***Caddoan: 3,500 BP?, 4,743 BP***		***-ka(ʔ)* GM**	**Swadesh 1958, ASJP, R: Taylor 1963**
	Pawnee (South Band)		*atí-ka* my GM	Taylor 1963
	Arikara		*atí-kaʔ* my GM	Taylor 1963
	Caddo		*i-káʔ* my GM	Taylor 1963
	Kitsai		*i’kani* GM	Bucca & Lesser 1969
	Wichita	*ŏkw* GF	*ŏ’kw* GM	Spier 1924
	Penutian stock: 5,522 BP?	**#*kaka* MB, MF, FZH**		**ASJP, R: AM&PB**
8	***Chinookan***			
	Chinook	*-gaga* MF, *qācqāc* FF		Boas 1904
	Tfalatik		*kaka* A	Ruhlen 1994
9	***Coosan***			
	Coos	*aχaax* MB		Sapir & Swadesh 1953

	Stocks, families	**Masculine denotata**	**Feminine denotata**	**References**
10	***Plateau***			
	Klamath-Modoc (Lutuami)	*ka˚ʐ-ip* MF, *kakas-ip* DS		De Angulo & Freeland 1931
	Sahaptian			
	Northern-Sahaptin	***táqaʔ* MB**		**R: Aoki 1966**
	Yakama	*káka* MB, *nakákas* my MB		Beavert & Hargus 2010
	Klikitat	*ka'ka* ad. MB, *naka'kas* my MB		Jacobs 1932
	Taidnapam	*ka'ka* ad. MB, *naka'kas* my MB		Jacobs 1932
	Palus	*tə'x̥a* ad. MB, *natə'x̥as* my MB		Jacobs 1932
	Walla-Walla	*tə'x̥a* ad. MB, *natə'x̥as* my MB		Jacobs 1932
	Nez Perce	*naʔtáx̥* my MB		Aoki 1994
	Yok-Utian: 6,500 BP?, 4,413 BP?			**Callaghan 1997, ASJP**
11	***Utian: 4,500 BP***			**Callaghan 1997**
	Miwokan 3,000 ~ 2,500 BP, 2,141 BP?	***kaka* MB, MBS, MBSS, etc., *#kawu* FZH**		**Callaghan 1997, ASJP, R: Callaghan 1997, AM&PB**
	Southern Sierra Miwok (Pohonichi)	*kaka* MB, MBS, MBSS, *kawu*, *kawy* FZH, ZH		Gifford 1922, Broadbent 1964
	Central Sierra Miwok	*ka'ka* MB, MBS, *ká˚wy* FZH, ZH, BinL		Gifford 1916, Freeland & Broadbent 1960
	Central Sierra Big Creek	*kaka* MB, *kawu* FZH		Gifford 1916
	Northern Sierra Miwok	*kaka* MB, MBS, FZH, *kawu* FZH, ZH, GDH, (♀)BDH		Gifford 1922, Callaghan 1987
	Plains Miwok	*kaka* MB, FZH, MZH, GMB, *kakatci* MBS		Gifford 1922, Callaghan 1984
	Coast Miwok	*kaka* MB, MBS, *gau* ZH		Gifford 1922
	Bodega Miwok	*káaka* MB		Callaghan 1970
	Lake Miwok	*kaka* MB, MBS, *kauko* FZH, ZH, ZHB		Gifford 1922
	Costanoan			
	Costanoan (Santa Cruz)	*howo* FF	*okko* GM	Pinart 1878, Gifford 1922
	Mutsun		*ka* D	Gifford 1922
12	***Yokutsan: 1,500 BP***	***ʔakas* MB (Proto-Nim-Yokut)**		**Callaghan 2001, R: Callaghan 1997**
	Paleuyami	*kokwat* MB, *kohotep* S, HBS		Gifford 1922
	Yaudanchi	*akash* ~ *agash* MB	*kawaiyis* MBW	Kroeber 1917, Gifford 1922

	Stocks, families	**Masculine denotata**	**Feminine denotata**	**References**
	Yauelmani	*agas ~ akash ~ ʔaagas̥* MB		Kroeber 1917, Gifford 1922, Sapir & Swadesh 1953
	Tachi	*agas* MB, MBS		Gifford 1922
	Gashowu	*n-u'ho* MB, MBS, MBSS		Gifford 1922
	Chukchansi	*n-o'ho* MB, MBS, MBSS		Gifford 1922
13	***Wintuan: 2,500 BP***	****k'iye ~ *k'iy* old man (GF, MB, FZH?)**		**Whisler 1980, Shepherd 2006, R: id. & AM&PB**
	Northwestern Wintun	*kiye* MF, MB, FZH		Gifford 1922
	Northern Wintun	*kiye* MB, MeBS, FeZS, *kiye-tcibet* P(♂) ancestor		Gifford 1922
	Northeastern Wintun	*kiye* MB, (♂)MBS		Gifford 1922
	Hayfork Wintu	*kiyeh* GF, MB, FZH		Whistler 1980
	Upper Sacramento Wintu	*kiyeh* GF, MB, FZH		Whistler 1980
	McCloud Wintu	*kiyeh* GF, MB, FZH		Whistler 1980
	Nomlaki	*kiyah* U?		Whistler 1980
	Paskenta Hill Nomlaki	*k'iya(h)* U (FB)		Whistler 1980
	Hill Patwin	*čiya(.)k* old man		Whistler 1980
	Southern Patwin	*čiy* old man		Whistler 1980
14	***Maiduan: 1,000 BP?, 1,219 BP?***	**#*kaka* MB, FZH** (or borrowed from Miwok?)		**ASJP, R: AM&PB**
	Northwestern Plains Maidu		*ka* yZ	Gifford 1922
	Northwestern Mountains Maidu	*ka* MB, FZH, M(♂)cousin	*gam* yZ	Gifford 1922
	Northeastern Maidu		*kam* yZ	Gifford 1922
	Southern Maidu	*kaka* MB, FZH	*ka* yZ, GGD	Gifford 1922
15	***Yuki-Wappo: P-Yukian 5,000 BP? Yukian***			**Campbell 1997**
	Coast Yuki	*kaha'* MB		Gifford 1922
	Yuki	*kikan* MyB, FyZH, *k'un* F, *kaï*^{n}*nt* FB	*k'an* M	Kroeber 1917, Gifford 1922
	Huchnom	*uñ-keka* MyB, FyZH, *ikika* FB, MZH		Gifford 1922
	Wappo	*awa* MeB, *ek'a* (♂)eBS, (♀)eZD		Gifford 1922
16	***Pailahnihan***			
	Achomawi	*akun* MF, (♂)DS	*akun* (♂)DD	Gifford 1922
	Atsugewi	*aqón* MF, (♂)DS	*aqón* (♂)DD	Gifford 1922, Garth 1944
	Hokan stock: 8,000 BP, 4,915 BP			**Golla 2007, ASJP**
17	***Washo***			
	Washo	*-gu* (♀)DS, *koi* F	*-gu* MM, (♀)DD	Kroeber 1907, 1917

	Stocks, families	**Masculine denotata**	**Feminine denotata**	**References**
18	***Pomoan***		**$^{\#}$*kats* MM**	**R: AM&PB**
	Eastern Pomo (Clear Lake)	*keha* ad. FB, stepF	*ghats* MM	Kroeber 1917
	Southeastern Pomo		*im-ka'* MM	Gifford 1922
	Northeastern Pomo		*katci-daiʔ* MM	Gifford 1922
	Central Pomo	*kegu* C, FZC, HBC, *eku* yB	*kate* MM, *eku* yZ	Gifford 1922
	Northern Pomo	*agi'* eB, *keghawi* ref. S	*kadai* ad. MM, *aka* ref. HM	Gifford 1922
	Southwestern Pomo		*kakan* MM	Gifford 1922
	Southern Pomo		*akatsen* MM	Gifford 1922
	Salinan			
	San Antonio	*kāi* eB		Mason 1912
	San Miguel	*kāiyE'* eB		Mason 1912
19	***Yuman-Cochimi***			
	Cocopa	*eny-ikwa* MB, *in-kwo* MF, *axga* DS, *agas* (♂)SibGS, *a'kas* (♀)SibGS	*ini-ka* MM, *agas* (♂)SibGD, *si'kas* (♀)SibGD	Gifford 1922
	Southern Diegueño	*iny-ikwai* MB, FZH, *in-kwau* MF, *axgau* (♂)DS	*in-kûs* MM, *axgau* (♂)DD	Gifford 1922
	Northern Diegueño	*in-ikwai* MB, FZH, *in-ikwa* MF, *ku'kau* SinL, *akatc* WeZS	*akatc* WeZD	Gifford 1922
	Mohave	*n-akwi-k* MB, *na-kweu-k* MF, *ahko'o-k* (♀)DC, *ahkyo'-k* (♂)DC	*n-akau-k* MM	Kroeber 1917
	Yuma	*n-ukwi* MB, *n-akwiau* MF, *axgo* DC	*n-ekyu* MM	Gifford 1922
	Kamia	*in-kwai* MB, *in-kwau* MF, *axgau* DS	*in-kas* MM, *axgau* DD	Gifford 1922
	Hualapai	*ko:* DC, *áwa* SC, *gwáwa* MF	*ko:* DC, *áwa* SC	Watahomigie *et al.* 2001
20	***Chumashan***			
	Chumash (Ynezeño)	*ma-qòqo* my F	*k-a-wa* my A	Gifford 1922
	Chumash (Barbareño)	*kó-ko* my F		Gifford 1922
	Chumash (Island)	*u-kâ'-kâ* my F		Gifford 1922
21	***Kiowa-Tanoan: 3,434 BP***			**ASJP**
	Taos		*k*^{l}*a-* M, EM, *k*^{l}*aʔù* M (diminutive)	Trager 1943
	Hopi-Tewa	*kuk'û˚* MF, *kóʔô˚ʔè˚* eZS	*kóʔô˚* my Z, *kóʔô˚ʔè˚* eZD, *kà˚káh* eZ	Dozier 1954
	Kiowa	*q'ų* ad.(♂)GF, (♂)GS	*qo'* ad. M, MZ	Lowie 1923
	Isleta	*k'*^{l}*aʔa* my F	*k*^{l}*iwʔù* ad. FZ	Trager 1943
22	***Comecrudan***			
	Comecrudo	*kia'm* U	*ke'm ~ ken* A	Swanton 1940

	Stocks, families	**Masculine denotata**	**Feminine denotata**	**References**
23	***Keresan***			
	Zia	*sawa'a* (♀)MB, *ano'wa* (♂)MB	*sa'ko'ye* (♂)Z	Hawley 1950
	Santa Ana	*sawá'a* (♀)MB, *sa'nawe* (♂)MB	*sakoi'tcʳ* (♂)Z	Hawley 1950
	Cochiti	*anawa* (♂)MB		Hawley 1950
	San Felipe	*s'anawá* (♂)MB		Hawley 1950
	Laguna	*s'anawe* (♂)MB	*ak'wi* (♂)Z	Hawley 1950
	Acoma	*sanawe* (♂)MB	*sˈakuich* (♂)Z	Hawley 1950
24	***Shastan***			
	Shasta	*kwako* (♀)B, *akwi* S		Gifford 1922
25	***Muskogean***			
	Koasati		*ki* M	Swanton 1928
	Creek	*łaha* eB	*tcki* M	Swanton 1928
26	***Uto-Aztecan: 8,900 BP?, 5,000 BP?, 4,100 BP?***	***_ka'_ FF, (♂)SS, *_kwaʔa_ MF, (♂)DCS, Southern Uto-Aztecan: *_ʔooka_ old woman**	***_ka'_ FM, (♀)SD, *_ko_ eZ**	**Merrill _et al._ 2009, Campbell 1997, ASJP, R: Voegelin, Voegelin & Hale 1962, Miller 1967, Wichmann 1999**
	Numic		***_kaku_** MM, (WM), (♀)DC	**Hage _et al._ 2004**[43]
	Kaibab Paiute		*kahu* MM	Kroeber 1917
	Battle Mountain Shoshone	*guno* FF, (♂)SS, *gago* (♀)DS	*guno* (♂)SD, *gago* MM, (♀)DD	Ives 1998
	Elko Shoshone	*kaku* (♀)DS, *kynu* FF, (♂)SS	*kaku* MM (♀)DD, *kynu* (♂)SD	Steward 1938
	Bannock	*igönu'* FF, (♂)SS	*igönu'* (♂)SD	Lowie 1930
	Piaviotso	*qönu'ᵘ* FF, (♂)SS	*qönu'ᵘ* (♂)SD	Lowie 1930
	Northeastern Mono	*gunu'* FF, (♂)SS	*gunu'* (♂)SD	Gifford 1922
	Southeastern Mono	*gunu* FF, (♂)SS	*gunu* (♂)SD	Gifford 1922
	Western Mono	*gunu* FF, (♂)SS	*gunu* (♂)SD	Gifford 1922
	Comanche	*ka'ku* (♀)DS, *kønu* FF, (♂)SS	*ka'ku* MM, (♀)DD, *nɜməkagu/* WM, WFZ, *kønu* (♂)SD	Gladwin 1948, Ives 1998
	Kawaiisu	*kuno-ni* FF, *kagu-ni* (♀)DS, *ku:či-ni* yBS, *kugu-ni* FeB	*kagu-ni* MM, (♀)DD, *ku:či-ni* yBD	Kroeber 1917, Shimkin 1941
	Northern Paiute	*kenu'u* FF, (♂)SS	*kenu'u* (♂)SD	Kroeber 1917
	Southern Paiute	*kenu'u* FF, (♂)SS	*kaku-* MM, *kenu'u* (♂)SD	Miller 1967
	Uintah Ute	*qu:-ni* FeB, *könu* FF	*kagu* MM	Kroeber 1917, Swanton 1913, Shimkin 1941
	Tubatulabal			
	Tubatulabal	*aka* FF, *kumu* FeB	*kuči* eZ	Kroeber 1917, Shimkin 1941

---

[43] Nichols, who is one of the authors of Hage (2004), has reconstructed the Proto-Numic kinship terms in 2002: *Notes on Proto-Numic kinship reconstructions* (unpubl. ms.).

Stocks, families	**Masculine denotata**	**Feminine denotata**	**References**
Takic			
Kitanemuk	*kukin* FF, *kwadi* MF, (♂)DS, *kwum* FeB	*kukin* FM, *kwadi* (♂)DD, *kor* eZ	Gifford 1922, Shimkin 1941
Varohio	*kukuri* FB, MF	*koči* eZ	Shimkin 1941, Miller 1967
Serrano	*kuki?* GGF, *-ka'* FF, SS, *kʷat* MF, (♂)DS, *kwa'* WF, *kumu* FeB, *aka* (♂)eBS	*kuki?* GGM, *-ka'* FM, SD, *kʷat* (♂)DD, *kakaiyek* (♀)PinL (after ego bears a child), *kakaiye* DinL (after she bears a child), *kër* eZ, *aka* (♂)eBD	Gifford 1922, Shimkin 1941, Miller 1967
Luiseño	*ka'* FF, (♀)EF, *ka'-mai* SS, *kwa?* MF, *kwa-mai* (♂)DS, *kwa-pa-na* HF, (♂)DH	*ka'* FM, (♂)DinL, *ka'-mai* SD, (♀)ZSD, *kwa-mai* (♂)DD	Kroeber 1917, Gifford 1922, Miller 1967
Cahuilla	*-ka'* FF, (♀)SS, *kum* FeB	*-ka'* FM, (♀)SD, *né-qa?* my GM, *kis* eZ	Shimkin 1941, Miller 1967
Desert Cahuilla	*ka'* FF, *kala* SS, *kwa* MF, *kwala* (♂)DS, *kux* GMB, *kuxhum* (♀)BGS, (♂)ZGS, *kum* FeB	*ka'* FM, *kala* SD, *kwala* (♂)DD, *kux* GFZ, *kuxhum* (♀)BGD, (♂)ZGD	Gifford 1922
Cupeño	*ka'* FF, *kama* SS, *kwa* MF, *kwana* (♂)DS, *kük* GMB, *kükima* (♀)BGS, (♂)ZGS, *kum* FeB	*ka'* FM, *kama* SD, *kwana* (♂)DD, *kük* GFZ, *kükima* (♀)BGD, (♂)ZGD	Gifford 1922
Hopi			
Hopi	*kʷa?a* MF, FF (♂)DS	*kʷa?a* (♂)DD, *ka?a* FZ, *iëqëqa'* eZ	Shimkin 1941, Miller 1967
Pimic			
Pima	*ka'k* (♀)SS, *kële* FeB	*ka'k* FM (♀)SD, *ka'kâks* FMZ, *hu?ul* MM, *âkâks* FeZ	Parsons 1928, Shimkin 1941
Papago	*ka?amad* (♀)SS	*kaak* ~ *ka?a* FM, *ka?amad* (♀)SD	Shimkin 1941, Miller 1967
Tepecano	*kaka-ri* (♀)SS	*kaka-ri* FM, (♀)SD	Shimkin 1941
Tepehuan	*boscica* [boski-ka] FF, *kumuli* FeB, *kasuli* (♀)SS	*kasuli* FM, (♀)SD, *kamata* id. *mat-kasuli* MinL	Radin 1931, Shimkin 1941
Tepehua - Santa Maria Ocotán (Southeastern)	*kuuxi'* MeB, (♂)eZS, *kuulsi* GF, (♂)GS, *o'kix* *(♀)*yZS	*kuuxi'* (♂)eZD, *kuulsi* (♂)GD, *o'kix (♀)*yZD, MeZ	Willett 2006-2010
Taracahitic			
Tarahumara	*acíkari* FF	*akácuri* FM, *cochi* [koci] eZ	Radin 1931
Opata	*kuh* FeB	*coo* [koo] ~ *ku* eZ	Radin 1931, Shimkin 1941
Corachol			
Cora	*-yaxu* FF, (♂)GS, *ne-yeé-k'wari* (♀) my GS	*-yaxu* (♂)GD, *ne-yeé-k'wari* my GM, (♀)GD, *-kuci* eZ	Radin 1931, Shimkin 1941
Huichol		*ne-kʷə́əci* my GM	Miller 1967
Aztecan			
Classical Nahuatl	*coli* ~ *colli* ~ *cohcolli* [ko7ko(l)li] GF, *ačka* (♂)eB		Molina 1555, 1571, Gardner n.d.

	Stocks, families	**Masculine denotata**	**Feminine denotata**	**References**
	Nahualt (Xalitla, Guerrero)	*-kohkol* GF, ancestor		Karttunen 1983-1992
	Mecayapan Nahuatl, Tatahuicapan de Juárez	*no-coco* [no-koko] my U, my eB, my eHB, *te-coco* [te-koko] masculine elder, *i-cójcol* [kohkol?] godfather		Wolgemuth, Walters *et al.* 2002-2010
	Tetelcingo Nahuatl	*i-cultzi* his GF		Brewer & Brewer 1971
	Cuisnahuat Pipil	*kuhkul* old man, *tu kuhkul* our GF		Campbell 1985
	Comasagua Pipil	*kuhkul* old		Campbell 1985
	Nicarao	*kuhkul* old man, evil spirit		Lara-Martínez & McCallister
	Isolates			
27	Tunica	*ki* MB, *ōka* S	*ōka* D	Swanton 1919
28	Chitimacha	*waʔ* MB, GMB, *kenʔ* FB, GFB	*koʔ* A	Swanton 1919, Haas 1939
29	Atakapa	*waxc* MB, FB, *hacka* eB, *hican* EF		Swanton 1919
30	Alsea	*hāʔt* eB		Ruhlen 1994b
31	Kootenay	*xa(ʔ)* FB, *x̥a’tsa* MB	*kukt'* MZ, MBW	Boas 1919, Morgan 1969
32	Coahuiltec	*ku-anāx* MeB, *ku-t’ān* MyB, *ku-ān* MF, *ku-ant’ān* GC, *kōu* FeB, *k’āu* H	*caca* [kaka] ~ *kaaka* FM	Garcia 1760 in Romney 1967, Swanton 1940
33	Haida	*qā’* MB		Mayer-Durlach 1929
	Haida Massett	*q'a* ref. MB, *ga’gai* ad. id.		Murdock 1934
	Haida Hydaburg	*q'a* ref. MB, *ga’ge* ad. id.		Murdock 1934
	Haida Skidegate	*q'a’-g̊a* ref. MB, *ha’g̊ai* ad. id.		Murdock 1934
34	Karok	*hogam* MB, (♂)ZS, *aka* F, *katc* S		Gifford 1922
35	Natchez	*gaga* (♂)eB		Swanton 1928
36	Seri	*axáac* MB (little used term)		Moser & Marlett 1997
37	Takelma		*xaga-* MZ, MBD, *xagā* ad. id.	Sapir 1907
38	Zuni	*kaka* MB, MMZS, HMB	*kuku* FZ, FZD, FZDD, HFZ	Schneider & Roberts 1956
39	Tonkawa		*ekak* ~ *ekac* GM	Sapir 1989

5. Synthesis of reconstructions

The following table offers a synthesis of the results obtained thus far, including some additional genetic units not developed in the present paper for reasons of space.

Stock and Region	Approx. time depth	Term	Meanings	Source of reconstructions	Proposed homeland
North America					
Proto-Inuit		*akkak	FB	Fortescue *et al.* 1994	
Proto-Yupik		*áka	eZ	EHL	
Proto-Athapaskan	2,000 BP	*ankayʔ	MZ	Hoijer 1956	Upper Yukon River (Golla 2007)
Proto-Algic	7,200 BP, 4,000 BP	#-oko	GM	AM & PB	Columbia Plateau (Golla 2007), NY State – CT (Wichmann 2010)
Proto-Algonquian	3,000 BP	*-okko-	GM, (EM)	Sapir 1922	Between Lakes Huron and Ontario (Siebert 1967), Columbia Pl. (Denny 1991)
Proto-Salishan	3,800 BP	*qá-	eB	Kinkade 1992	Between Skagit & Fraser Rivers (Kinkade 1991)
Proto-Salishan	3,800 BP	#sxaxa	EF, DH, SW	Morgan 1980	*idem*
Proto-Siouan	3,000 BP	*khų	GM, EM	Matthews 1959	Ohio valley?
Proto-Caddoan	3,500 BP	*-ka(ʔ)	GM	Taylor 1963	Southern Arkansas (Wichmann 2010)
Proto-Penutian	5,500 BP	#kaka	MB, FZH	AM & PB	North Great Basin, Columbia Pl. (Golla 2007)
Proto-Sahaptian		*táqaʔ	MB	Aoki 1966	Columb. Pl. (Kinkade 1991)
Proto-Miwokan	2,000 ~ 3,000 BP	*kaka	MB, MBS	Callaghan 1997	North of San Francisco Bay (Callaghan 1997)
Proto-Nim-Yokuts	1,500 BP	*ʔa-kas	MB	Callaghan 2001	Pre-Proto-Yokuts in Great Basin? (Golla 2007)
Proto-Wintuan	2,500 BP	*k'iye	old man, (MB)	Whistler 1980	Interior NW California – SW Oregon (Whistler 1977), W Oregon (Golla 2007)
Proto-Maiduan?	1,000 BP	#kaka	MB, FZH	AM & PB	Pre-Proto-Wintuan: Northern Great Basin (Golla 2007)
Proto-Yukian-Wappo	5,000 BP	#keka	MB, FZH	AM & PB	Eel River drainage, NW California (Foster 1996: 83)
Meso-America					
Proto-Totonac-Tepehua	2,400 BP	#koko ~ kuku	U (MB)	AM & PB	Builders of Teotihuacan (AD 200-650: Campbell 1997)
Proto-Jicaque (Tol)		*kok^{h}am *(ko)koy *(ku)kus	U (♂)GF D	Campbell & Oltrogge 1980	
Proto-Mixe-Zoquean	3,000 BP	*-oko	GM, GC	Wichmann 1999	Spoken by the Olmecs (Campbell 1997), Tuxtla Mountains?
Proto-Mayan	4,000 BP	*ikaan *ikaaq	(♂)MB, (♂)FZH, (♂)MGF, WF cousin	Wichmann & Brown n.d., Kaufman & Justeson 2003	Cuchumatanes Mountains, Guatemala (Campbell 1997)
Proto-Uto-Aztecan	8,900 BP ~ 5,000 BP	*ka' *kwaʔa	FF, FM, (♂)SC MF, (♀)DC	Miller 1967, Voegelin,	Great Basin USA (Merrill *et al.* 2009), Arizona – northern Mexico (Fowler

Stock and Region	Approx. time depth	Term	Meanings	Source of reconstructions	Proposed homeland
				Voegelin & Hale 1962	1983), Meso-America (Hill 2010)
Proto-Otomanguean	6,500 BP ~ 4,000 BP	*_hkeh_ *_kha_ *_kuHn_ *_kwaHn_	GF, GS male affinal kin sibling (♀)eZ	Merrifield 1981	Tehuacán Valley (Hopkins 1984, Campbell 1997)
Proto-Chibchan	6,700 BP	*_kaka_ *_gaka_	F EM	Constenla Umaña 1981	South Central America (Constenla Umaña 2012)
Proto-Magdalenic	5,200 BP	#_kaka_	GM	AM & PB	Northern Colombia?
Proto-Lenca-Misumalpan	7,200 BP	#_kVkV_	GF, GM, anciano/a	Constenla Umaña 2002	Honduras – El Salvador (Constenla Umaña 2002)
South America					
Proto-Cariban	3,700 BP	*_koko_	U	De Goeje 1946	Venezuelan Guiana (Villalón 1991)
Proto-Cariban	3,700 BP	#_kuku_	GM, FZ, EM	AM & PB	*idem*
Proto-Taranoan (Cariban)	500-900 BP	*_kuku_	GM (FZ)	Meira 1998	Central & Southern Surinam (Meira 1998)
Proto-Pano-Takanan	4,700 BP	#_kuku_ ~ #_kuka_	MB, EF	Girard 1971	
Proto-Panoan	2,000 BP?	*_koka_	MB [EF]	Girard 1971	East Central Peru (Fleck 2013)
Proto-Panoan	2,000 BP?	*_koko_	nephew	Girard 1971	*idem*
Proto-Takanan	2,000 BP?	*_kuku_	U (MB)	Girard 1971	Northern Bolivia?
Proto-Arawakan	4,500 BP	*_kuko_ *_aku-_	U, EF A	Payne 1991	Middle Orinoco – Upper Amazon (Heckenberger 2002)
Proto-Arawan	1,700 BP	*_koko_	ad. MB, EF	Dixon 2004b	Middle Juruá River?
Proto-Guahiboan	2,300 BP	*_-axu_ *_akwe_	MB, EF GM	AM & PB	Northern Colombia (Wichmann n.d.)
Proto-Quechuan	> 1,500 BP	*_kaka-_	MB, EF	AM & PB	Central Peru (Adelaar 2012)
Proto-Aymaran?	> 1,500 BP	#_kaka_	MB, EF	AM & PB	Central Peru (Adelaar 2012)
Proto-Nambikuáran	2,800 BP	#_kũka-_	GF, MB, EF	AM & PB	Upper Juruena River (Wichmann n.d.)
Proto-Karirían	> 300 BP	#_kuku(h)_	MB	AM & PB	Lower Rio São Francisco - State of Paraíba (Brazil)?
Proto-Mondé	2,000 BP	#_koko_	ad. MB, EF	AM & PB	Roosevelt & Aripuana Rivers (Brunelli 1987), Jiparana River headwaters (Campbell 1997)
Proto-Kaingang	?	*_kakra_ *_kake_	MB, EF eB	Jolkesky 2010	Parana State (Brazil)?

Table 13. _KOKO_ forms in various proto-languages of America. Proto-forms preceded by an asterisk * have been reconstructed by specialists of the concerned group, as indicated in column Sources; proto-forms preceded by a hash mark # are not reconstructions, but are postulated by us (AM & PB) according to the generalized presence, in all or nearly all member languages of the concerned group, of forms

Stock and Region	Approx. time depth	Term	Meanings	Source of reconstructions	Proposed homeland

either identical to the putative etymon or clearly derived from it. References for the ages of stocks have already been given in tables 11 and 13 and in section 2.

The time depths reported in our paper have been calculated either using the method of glottochronology or the automated dating method based on lexical similarity used by ASJP. Although the time depths obtained by these methods differ somehow, all of them are to be counted in thousands of years. By coupling time depths of language families with reconstructed *KOKO* terms, table 13 makes apparent that *KOKO* terms were in use several millennia ago by the ancestor languages of many "major" linguistic families (major in terms of the number of languages they include), whose postulated homelands were, most of the time, far away from each other (see maps 1, 2, 3). Among the most ancient stocks or families, let us mention Lenca-Misumalpan (7,200 BP), Chibchan (6,700 BP ~ 4,484 BP), Algic (> 3,000 BP), Arawakan (Maipuran) (4,500 BP ~ 4,134 BP), Pano-Takanan (4,700 BP), Cariban (3,700 BP ~ 2,362 BP), Uto-Aztecan (8,900 BP ~ 5,000 BP ~ 4,118 BP), Yok-Utian (5,000 BP ~ 4,413 BP), Caddoan (3,500 BP ~ 4,828 BP), Siouan (3,000 BP ~ 3169 BP), Oto-Manguean (6,500 BP ~ 7,418 BP). A *KOKO* form has not been posited at the proto-level of some other major stocks like Jêan (4,989 BP), Tupían (5,500 BP ~ 3,585 BP), Tukanoan (2,699 BP), Siouan-Catawban (4,000 BP ~ 6,856 BP), although some families or branches belonging to these groups do display *KOKO* terms, sometimes even at the level of the proto-language, like Mondé (2,000 BP) for the Tupían family, Kaingang (no date) for the Jêan family, Siouan (2,500 BP ~ 3169 BP) and Catawban for the Siouan-Catawban stock. *KOKO* forms have been reconstructed for "younger" families like Quechuan (1,500 BP ~ 1,717 BP) or Arawan (1,764 BP), and are also postulated in some of the "smaller" families like Nambikuáran (2,807 BP), or Karirían. Finally, as we already mentioned, they are also present in 60% of the language isolates for which there are substantial data (maps 1 and 2: South and Meso-America).

- ***Semantic scope***

Table 14. Semantic distribution of *koko* forms in proto-languages and isolates

Meanings	GF	MB	Sib	GM	Other
Total = 56	5	25	4	11	11
Percentage	8.9%	44.6%	7.1%	19.6%	19.6%

At this stage of our study, we will just give a general idea of the semantic scope of the form *KOKO* at the linguistic family level, drawn from table 13. A more fine-grained analysis of the semantic scope will be performed in the anthropological part of our paper. For now, the meanings MB/EF, MB/FZH, MB/MBS are rallied in the category [MB], so much for the meanings GM/EM and GM/FZ which are rallied in the category [GM]. This results in the figures given in Table 14.

6. Discussion: the origin of *KOKO* in the Americas

- ***A matter of convergence?***

Is it possible that the continental distribution of *KOKO* terms should be the result of chance or of some convergence global process? We will refer the reader to some of our previous publications and notably Bancel & Matthey (2002) in which this question was thoroughly treated at the world level for the etymon *KAKA*. We will also refer them to Matthey & Bancel (2008), Matthey, Bancel & Ruhlen (2011), Bancel & Matthey (2013), showing that contrary to Trask's (2004) allegations, nursery kinship terms are not perpetually recreated or reinvented in the world's languages, as former nursery terms get linguistically eroded, but are deeply rooted in linguistic families, and transmitted faithfully from one generation to another, through millennia, with very few phonetic transformations. As we stressed in Bancel & Matthey (2002), and the following publications, the daily use of this term as well as the use of other reduplicated terms like *PAPA* and *MAMA*, generally called "nursery terms", the ease of their transmission to young children, their high symbolic significance, have made them extremely resistant to phonetic and semantic change, as is otherwise fully demonstrated by their continuous written transcriptions in the course of the past 5,000 years or so (Matthey de l'Etang & Bancel 2008, Bancel & Matthey de l'Etang 2013). Likewise, *KAKA* terms, just like their *PAPA*, *MAMA* and *TATA* counterparts, have also left traces in the written records of Indo-Hittite languages enabling the reconstruction of Proto-Indo-Hittite **HawH-os (*xawx-os)* MB, GF (Nikolayev 2007), likewise in the written records of Chinese since Preclassic Old Chinese *guʔ* MB some 3,000 years ago (Starostin 2005). But even with no ancient written records left behind, there is little doubt that $^{\#}$*ka(a)ka* GP in Niger-Congo is extremely ancient if not the proto-form in this language phylum. This very same form **-kààká* GP has been reconstructed in Proto-Bantu by both Meeussen (1969) and Guthrie (1967-1971).

Let us add that, in our opinion, the hypothesis that the transcontinental distribution of the same forms and meanings, let us say *KOKO* MB, in the languages of the Americas results from sheer coïncidence appears utterly improbable. Unless some underlying principle attaching CVCV reduplicated forms, with velar consonant, in occurrence *KOKO*, to the MB or GM relationships can be demonstrated, the convergence theory would predict forms with more erratic meanings, than just those, found in all regions, and for the most part consistent with Dravidian-type terminology.

➢ *Diffusion and borrowing*

The second hypothesis that must be debated is the possibility that the general distribution of *KOKO* terms all over the Americas results from an ancient but still ongoing diffusion-process, making this region as a whole look like a linguistic area.

Upholders of such an hypothesis must explain, and provide some kind of evidence as to how cascade borrowing accounts for the phonetic and semantic parallelisms of the American series *KOKO*, the huge accumulated evidence of its transcontinental distribution, and the apparent linguistic retention of *KOKO* forms, as shown in their reconstructions in linguistic families, whose supposed time depths are sometimes as remote as 7,000 BP.

It is worth mentioning at this point that our survey of kin terms diffusion among intermarrying language groups, including those practicing linguistic exogamy[44], has not provided support for a model of mass diffusion of kinship terms, as the one envisioned by Dixon & Aikhenvald (1999: 8; see Presentation), notably concerning *KOKO* forms (Matthey de l'Etang & Bancel in preparation). Our survey of the ethnological literature includes the Arawakan-Tukano-Maku cluster in the Vaupés basin[45], the Arawakan-Tupi-Cariban-Trumai cluster of the Upper Xingu region, the Arawakan-Cariban cluster of the Western Indies, as well as the relationships between Arawakan and Takanan in Bolivia, between Arawakan and Panoan in Eastern Central Peru, between Quechuan and Aymaran in Peru, and between Arawakan and Quechuan in Peru. The explanation given by these authors does not match the facts correctly, notably at the semantic level, as *KOKO* terms do substantially refer to kin types different from just EF/MB, and particularly to the GM. This casts a serious doubt on the reality of the borrowing mechanism hypothesized by Dixon & Aikhenvald.

---

[44] "Linguistic exogamy" refers to marriage prescribed between groups speaking different languages, and implies that one spouse will join the residence of the other (see details in notes 45 and 46).

[45] In the Vaupés region (north-east Brazil), since residence is patrilocal, a woman has to leave her paternal group and join her husband's. Social identity is *"established by patrilineal descent and has language group affiliation as its primary marker;"* identification is done with *"one's father's language group"* (Stenzel 2005: 3-4). As Sorensen (1967: 677) puts it, *"an individual belongs to his (or her) father's tribe, and to his father linguistic group, which is also his own."*

This is not to say that a number of divergent phonetic forms of *KOKO* within a language family series do not result from the borrowing of forms from other families. But most of the documented cases of linguistic and ethnic contact show that kinship terms are seldom borrowed by one group from another, because language appears determinant in keeping the group's identity.[46]

But our key contention regarding the theory of mass diffusion does not basically rest on an empirical demonstration, but on a theoretically compelling argument, recently developed by Martin Haspelmath facing the problem of knowing whether Australia as a whole could actually be defined as a linguistic area. According to Haspelmath (2004: 211), *"linguistic areas need not only be internally coherent, but also distinctive with respect to languages outside the area. Thus, one would have to show that the Australianisms are uncommon in the rest of the world, or at least in adjacent areas."* Transposed in the particular case of the American *KOKO ~ KAKA* MB, GF, FZ, GM, eSib, the second criterion is not met, because this word is in fact one of the most common kinship terms in the world, and is found in languages of New Guinea, Australia, Africa, Eurasia, Oceania, with similar meanings, and consequently does not represent a distinctive feature with respect to the languages outside the Americas.

We can then assume that the pan-American distribution of *KOKO ~ KAKA* MB, GF, FZ, GM, eSib does not result from the borrowing or diffusion of an American areal feature, but looks consistent with its inheritance from kinship terms present in the language or languages spoken by the group or groups which first colonized the Americas. This better explains why, as far back in time as comparative linguistics can point out, a number of language groups in their ancestral stage, as well as some ancestors of the languages isolates spoken across the Americas in areas very distant from one another, as maps 1, 2 and 3 above do show, had *KOKO* forms in their kinship lexicon. Thus, if we may venture a paraphrase of a famous comment made by Sapir about the distribution of first person root *n-* across the Americas: *"How in the hell are we gonna explain the general American KOKO except genetically?"*[47]

- **KOKO *and the peopling of the Americas***

Of course, the central question that one will ask at this point is: how do the massive geographic and linguistic distribution of *KOKO* kinship terms, and the fact that they have been preserved through millennia and transmitted with very little or no modifications within families at a high level, fit into the history of the peopling of the Americas?

---

[46] Stenzel (2005) speaking of the Vaupés plurilingual situation, Seki (1999) speaking of the Xingu situation, after Sorensen (1967), have stressed the role of language as a marker of identity, as one individual always associates himself with his (her) parent's language according to the line of descent.

[47] In a personal letter to Franck Speck, dated from August 1, 1918, Sapir wrote: "*Getting down to brass tacks, how in the Hell are you going to explain general American n- 'I' except genetically? It's disturbing, I know, but more non-committal conservatism is only dodging, after all, isn't it? Great simplifications are in store for us.*" (Quoted in Darnell and Hymes (1986: 229-230).

Following from Ruhlen (1994a & b), Bancel & Matthey de l'Etang, and Matthey de l'Etang & Bancel (2002, 2011), the massive distribution of *KOKO* kinship terms in America appears as a local development from the Proto-Sapiens etymon *KAKA* EF, MB, GP. It is the consequence of the colonization of the Americas by one or several groups, whose kinship terminologies originally comprised *KOKO* kinship terms. But there is certainly one thing that the distribution pattern does not tell us, in and of itself, and this is the number of language groups that were involved in the migration process. The global distribution of *KAKA* ~ *KOKO* words is consistent with both a single-migration and a multi-migration model.

One of the main contentions concerning the initial peopling of the Americas bears on the number of migration waves that occurred during the colonization process: three temporally distant waves according to the well-known and much debated conclusions of Greenberg *et al.* (1986)[48], only one according to many of the recent genetic studies of mtDNA, Y-chromosome, autosomes in American Native populations. These two conclusions may appear contradictory, but they are not, as we shall see.

It is noteworthy that the three language families defining Greenberg's migration waves, i.e. Amerind, Na-Dené and Eskimo-Aleut, all display *KOKO* terms to a certain extent, as is indicated in tables 12 and 13, a fact apparently in line with the observations made in the preceding paragraph. Thus, **aka* eZ, M has been reconstructed in Proto-Eskimo (EHL), **akkak* FB has been reconstructed in Proto-Inuit (Fortescue *et al.* 1994, EHL)[49], while no proto-*KOKO* form has been reconstructed for either Proto-Athapaskan[50], or Nuclear Na-Dené (Eyak-Tlingit-Athapaskan), even if Eyak and Tlingit do display such forms. Let us finally mention that Haida, which is considered a Na-Dené language by Greenberg and Ruhlen, does have *q'a'-g̊a* 'my MB'. All the other *KOKO* forms have been attributed to the Amerind macrophylum by Ruhlen.

➢ *The genetic studies of Native Americans*

1. *A major migration wave along the Pacific coast ca. 16,000 BP*

The most recent genetic studies of Native Americans have stressed the fact that virtually all Native American populations trace their ancestry to a limited number of founder mtDNA haplogroups, A2, B2, C1, D1, and D4[51], with similar coalescence times, and to the Y-chromosome haplogroups Q1a3a, Q1a3*, C3. These studies consequently make the assumption that only one population wave of Asiatic origin[52] was responsible for the initial peopling of the Americas, and for spreading the pan-American haplogroups into the continent (Tarazona-Santos & Santos 2002, Zegura *et al.* 2004, Tamm *et al.* 2007, Fagundes *et al.* 2008, Achilli *et al.* 2008, Kumar *et al.* 2011). In line with these conclusions, it has been shown that the very same genes were present in archaic human remains dated from 13,000 to 4,000 years BP, all over the Americas (Smith *et al.* 2005, Kemp *et al.* 2007, Manriquez *et al.* 2011, Chatters *et al.* 2014). Finally, Schroeder *et al.* (2007, 2009) have taken the high frequency of a private allele, *"the 9-repeat allele at microsatellite D9S1120 in all sampled Native American and Western Beringian populations... as evidence that all modern Native Americans descend primarily from a single founding population"* (2009: 995).

---

[48] Zegura *et al.* (2004: 164) write: *"In 1986, Greenberg, Turner & Zegura published a widely cited, synthetic, position paper on the early peopling of the Americas that stressed the apparent congruence of the then available data from linguistics, dental morphology, and traditional biparental nuclear genetic systems within the context of the archaeological record. Their major explanatory hypothesis, the 'three-wave' or 'tripartite' model, was based on the proposition that all indigenous Native American populations could be allocated to three distinct linguistically defined groups (i.e., Amerind, Na-Dené, and Aleut-Eskimo) that had their origins in three chronologically separate migrations from different geographic areas of Asia (Greenberg, Turner & Zegura 1986)."*

[49] Aleut refers to the GM as *kukaq*.

[50] Except perhaps *ank̔ay?* MZ in Proto-Athapaskan, reconstructed by Hoijer in 1956.

[51] Perego *et al.* 2009 identified the mtDNA D4h3 haplogroup in California, Mexico, Peru and Chile and suggest, based on its age estimate of ca. 16,000 BP, that it entered the Americas with the major wave along the Pacific coast.

[52] The sequencing of the genome of a ±24,000-year-old anatomically modern human individual from Mal'ta in south-central Siberia by Raghavan *et al.* (2014) revealed mtDNA and Y-chromosome

This (major) migration wave, generally dated back to ca. 18,000-14,000 BP – thus predating the Clovis culture (Zegura *et al.* 2004, Tamm *et al.* 2007, Fagundes *et al.* 2008, Achilli *et al.* 2008, Kumar *et al.* 2011) – apparently rapidly followed a Pacific coastal route, making its way into South America[53] (Tamm *et al.* 2007, Wang *et al.* 2007, Fagundes *et al.* 2008, Reich *et al.* 2012, Bodner *et al.* 2012). These dates are in accordance with those given in recent archeological studies of Paleo-Indian sites (Goodyear 2005). Besides, Wang *et al.* (2007: 2049), studying the genetic diversity and population structure in the Americas, observed *"gradients both of decreasing genetic diversity as a function of geographic distance from the Bering Straits and of decreasing similarity to Siberians – signals of the southward dispersal of populations from the northwestern tip of the Americas."* They concluded (2007: 2059) that this *"genomic continent-wide pattern"* is consistent with a model in which, *"at each step in the migration, a subset of the population splitting off from a parental group moves deeper into the Americas, taking with it a subset of the genetic variation present in the parental population."*

2. *The presence of other lineages*

Additionally, a few other minor lineages such as the mtDNA haplogroups X2a, D2, D3, C4c, or the Y-chromosome haplogroups Q1a5[54] and Q1a6 have been reported in Native American populations (Tamm *et al.* 2007: 4, Perego *et al.* 2009, Kumar *et al.* 2011, Dulik *et al.* 2012), but they appear to be restricted to North America.[55] The presence of these lineages raises the question

---

haplogroups, U and R respectively. MtDNA haplogroup U has been found at high frequency among Upper Palaeolithic and Mesolithic European hunters-gatherers, while the Y chromosome haplogroup R is basal to modern-day western Eurasians. The same study revealed autosomal evidence that the Mal'ta individual was basal to modern-day Eurasians, and also genetically closely related to modern-day Native Americans, with no close affinity to East Asians: 14% to 38% of Native American ancestry may originate through gene flow from this prehistoric population. What Raghavan *et al.* suggest is that this gene flow occurred *"after the divergence of Native American ancestors from east Asian ancestors, but before the diversification of Native American populations in the New World."* One of the implications of the study, according to the authors, is that the results *"may provide an explanation for the presence of mtDNA haplogroup X in Native Americans, which is related to Western Eurasians but not found in East Asian populations."* Another implication that the study provides is *"a possibility that the non-East Asian cranial characteristics of the First Americans derived from the Old World via migration through Beringia, rather than by a trans-Atlantic voyage from Iberia as proposed by the Solutrean hypothesis"* (Raghavan *et al.* 2014: 89, and see note 55).

[53] Less than 2,000 years for the entire Pacific coast, according to Bodner (2012: 6).

[54] Dulik *et al.* (2012: 2) report that one Tłįchǫ, one Slave and possibly one Alaskan Athapaskan belong to Q1a5.

[55] These clades have only been identified in North America. The haplogroup X (X2a) is found in North America at a low frequency. The fact that its coalescence time appears younger than those of the American haplogroups A-D has made some researchers argue that haplogroup X *"represents an*

as to whether they represent minor founding haplogroups participating in the same (major) population expansion or lineages belonging to more recent gene flows.

3. *X2a and C4c: an inland route?*

The phylogeographic analyses of the mtDNA haplogroup X2a, identified in Native populations of North America, notably in Algonquian, Wakashan, and Sahaptian speakers (Perego *et al.* 2009), and of the mtDNA haplogroup C4c detected in Cherokee, Creek, Siouan, Chippeway (Algonquian), Shuswap and a few other individuals from unknown ethnic origin (Kashani *et al.* 2012), whose coalescence ages appear similar to those of the pan-American mtDNA haplogroups, make the authors (a single research team) suggest that X2a and C4c could possibly have entered the Americas with a second major migration wave using the ice-free corridor between the Laurentide and Cordilleran ice sheets, more or less concomitantly to the first coastal migration-wave. The same research group (Achilli *et al.* 2013) wonders if this additional migration wave could also concern the Na-Dené speaking groups, because some of them – the southern Athapaskan, notably – do show the presence of X2a. Suggesting that X2a and C4c could actually be present in northern Na-Dené speaking groups as well, the same authors conclude (2013: 5) that *"the intermediate migration highlighted by nuclear data in the Chipewyan by Reich* et al. *would be part of a larger-scale migratory event that did not affect only the ancestor of Modern Na-Dené"* (see 7 below).

4. *Genes and Na-Dené language*

No one-to-one correspondence between genetic data and the Na-Dené-speaking population has been clearly demonstrated. The two mtDNA haplogroups reported at high frequencies in this group are in the first place a sublineage of the pan-American A2, i.e. A2a, and secondly a sublineage of the Siberian-Beringian D2, i.e., D2a1a,[56] whose age is estimated back to 6,900 BP ± 4,100y (Volodko *et al.* 2008). Both these haplogroups are also reported in Eskimo-Aleut and Chukchi populations. Volodko *et al.* (2008; 1087) report that the pan-Amerindian *"A2 mtDNA coding regions available from GenBank share no mutation"* with the Chukchi and Na-Dené A2a lineage, except for the A2 root, and conclude (2008; 1095) that the *"geographic specificity of these lineages confined to Chukotka and Alaska is the main argument in favor of the refugial hypothesis which assumes the origin of the founding populations of the Eskimo-Aleut and Na-Dené Indians in the southern Alaska at the terminal Pleistocene or early Holocene."* This conclusion was also reached by Starikovskaya *et al.* (2005): *"The geographic specificity and phylogeny of haplogroup D complete sequences support the refugial hypothesis which proposes that the founding populations of Eskimo-Aleuts and Na-Dené Indians originated in the eastern Beringian/Alaskan refuge area during the early postglacial period."*

Besides, studies of paternally inherited Y-chromosomal DNA has long revealed that Na-Dené populations notably harbor the Y-chromosome haplogroups Q1a3 and C3 that Zegura *et al.* (2004) also identified in Native American and Eskimo groups, making them endorse a one-wave

---

independent migration from Asia or even Europe" (Fagundes *et al.* 2008: 1, and references given for the "Solutrean" hypothesis). Perego *et al.* (2009) reported a coalescence time of ca. 16.7 to 15.5 ky for all the Native American clades, i.e. including X2a. D2 has been identified in Siberia and in North America (Na-Dené and Eskimo-Aleut populations), and D3 has been identified in Siberian and Eskimo populations.

[56] The presence of B2 in Athapaskan-Navajo populations is likely the result of admixture with their non-Athapaskan neighbors (Torroni *et al.* 1992).

migration model, all the more so since both haplogroups yielded rather similar coalescence time estimates at respectively 14,700 BP ± 5,700y and 13,600 BP ± 4,100y (lower bound) (Zegura *et al.* 2004). But in fact, there are two C subclades which have been identified in the Americas, C3b or C-P39 having the P39 marker, and a more ancient C3* or C-217 without the P39 marker. The first haplogroup has been identified in North America mainly in Athapaskan-speaking populations,[57] whereas the C3-M217 clade has been identified in a Tlingit-speaking individual from Southeast Alaska, in 8 Waorani and 11 Kichwa speakers from Ecuador and in 2 Wayuu-speaking individuals from Colombia (Roewer *et al.* 2013). The time to the most recent common ancestor (TMRCA) for C3b Y-chromosomes given by Dulik *et al.* (2012) is ca. 5,000 to 10,500 BP, depending on the statistical programs used, signaling, according to the same authors, a population expansion involving *"mostly Athapaskan speakers"* (2012: 5), different than that of the first (major) wave.[58] Let us remark that the coalescence age estimates of C3b correspond better to the age of the Na-Dené language family, ca. 8,000 BP according to AJSP, than to the age of Proto-Athapaskan, ca. 2,500 BP (Krauss 1973).

5. *Evidence for a Paleo-Eskimo migration ca. 5,500 BP*

The multiwave hypothesis has otherwise received support from the genomic sequencing of a permafrost-preserved hair belonging to a Paleo-Eskimo individual found in a Saqqaq culture context, in Qeqertusussuk on the west coast of Greenland. The archeological site, excavated between 1983 and 1993, dates back to 3,900-3100 yBP. The individual was assigned to the mtDNA haplogroup D2a1, which is closely related to the common mtDNA haplogroup reported notably in present-day Aleut populations of the Commander Islands (Gilbert *et al.* 2008: D2a1a; Volodko *et al.* 2008: D2a1a1), in some Siberian Sireniki Eskimos (Gilbert *et al.* 2008: D2a1b; Volodko *et al.* 2008: D2a1a), and in a number of Inupiaq-speaking individuals from the Alaska North Slope region (Raff *et al.* 2011: D2)[59]. In another paper dedicated to the same Paleo-Eskimo, Rasmussen *et al.* (2010) compared the high-confidence single-nucleotide polymorphisms (SNPs) of the Saqqaq Eskimo to those of contemporary populations in order to find which were the most closely related to this individual. The conclusions were that the Saqqaq individual was more closely related to present-day Nganasans, Koryaks and Chukchis from north-eastern Siberia than to the Amerinds, the Na-Dené speakers, or the Greenland Inuits. Thus the Saqqaq population apparently shared ancestry with Arctic East Asians and not with the people from the first migration wave. The genetic proximity between the Koryaks and the Saqqaq individual is also apparent in the assignment to the Y-chromosome haplogroup (or paragroup) Q1a*[60] of both the Saqqaq individual and the four Koryaks inhabiting the Sea of Okhotsk coast (Malyarchuk *et al.* 2011). Bisso-Machado *et al.* (2011), for their part, assigned one Siberian Yupik Eskimo to the same paragroup Q1a*.

Gilbert *et al.* (2008) and Rasmussen *et al.* (2010) concluded that these remains (if representative of a population) constitute evidence for a population expansion along the northern fringes of the American continent, different from that of the first wave, and also different from the Thule expansion that occurred ca. 1,000 years ago. This population wave is dated by Gilbert back to ca. 4,500 BP, and by Rasmussen *et al.* (2010: 757) to ca. 5,500 BP, based on the estimated mtDNA genetic divergence date between the Chukchis and the Saqqaq individual of between 4,400 and 6,400 years BP, coupled to the oldest archeological evidence of the Arctic Small Tool tradition dating back to 5,500 BP (Rasmussen *et al.* 2010: 761). Dulik *et al.* (2012) argued that this population expansion and that involving Athapaskan-speaking populations which was mentioned two paragraphs above are basically concomitant.[61]

---

[57] Also at moderate frequencies in Siouan and Cheyenne populations (Zegura *et al.* 2004).

[58] Dulik *et al.* (2012) reactivate a model supported by Lell *et al.* in 2002, and Bortolini *et al.* in 2003, but challenged by Tarazona-Santos & Santos (2002) and Zegura *et al.* (2004).

[59] See also Zlojutro 2006-2008, Crawford *et al.* 2010.

[60] The word *paragroup* refers to a lineage of a haplogroup not defined by any specific additional marker, and written with an asterisk *, here Q1a*.

Dulik *et al.* (2012) proposed to equate the Y-chromosome Q1a* of the Paleo-Eskimo man to the haplogroup Q1a6, which is defined by the NWT01 marker and whose coalescence age between 7,000 and 5,000 BP[62] is consistent with the age of the oldest evidence of the Arctic Small Tool tradition ca. 5,500 BP. Furthermore, given the fact that Q1a6 has been reported in Canadian Inuits (Inuvialuit), in Northern Alaskan Inupiaq-speaking communities, in some Yupik populations, and also inferred in the four Koryaks previously assigned to Q1a*, Dulik *et al.* (2012) assume a continuity between the Paleo-Eskimo individual and modern Inuits and Yupiks.

6. *The mtDNA haplogroup D3*

Finally, the mt-DNA haplogroup D3 is present in a number of Siberian populations such as the Nganasans, Yukaghirs, Chuvantsis and Chukchis, as well as in Siberian Naukan Eskimos (Volodko *et al.* 2008), some Aleut individuals (Crawford *et al.* 2010, Zlojutro 2006-2008) and in Alaskan, Canadian and Greenlandic Inuits (Helgason *et al.* 2006, Raff *et al.* 2011). This haplogroup is generally associated with the Thule expansion, dated close to ca. 1,000 AD (Gilbert *et al.* 2008).

7. *New perspectives*

Before we put a provisional full-stop to this section, let us mention that a multiple-wave scenario is also supported in a highly comprehensive survey of genetic diversity in Native Americans, carried out at the highest resolution level by Reich *et al.* (2012), showing that the Saqqaq individual, the Aleutian peoples, and the East and West Greenland Inuits derive 57% of their genetic ancestry from admixture with populations descending from the first migration wave, and 43% from specific Asian lineages, marking a distinct migration from Asia. The same study shows that the Athapaskan (Na-Dené) speaking Chipewyan inherit 90% of their genetic ancestry from likely admixture with populations descending from the first migration wave, and 10% from a third gene stream also marking a distinct migration from Asia. The paper furthermore indicates that the Asian lineages leading to the Eskimo-Aleuts on the one hand and the Na-Dené-speaking Chipewyan on the other hand, are closely related and apparently *"descend from a Siberian population that is a sister group to the Han."* Reich *et al.* (2012: 372) also stress the fact that they have data from just one Na-Dené-speaking group, and that *"an important direction for future work will be to test whether the distinct Asian ancestry that we detect in the Chipewyan is a shared signature throughout Na-Dené speakers."*[63]

---

[61] Achilli *et al.* (2013) associate the A2a clade of the Pan-American A2 mtDNA haplogroup, which is found among Athapaskan and Eskimo-Aleut speaking populations, and notably the variant found among the latter (A2a2 and A2a3) which *"experienced the steepest population expansion"* ca. 4,000 BP, to the Paleo-Eskimo wave.

[62] Dulik *et al.* (2012: 4) write: *"The Y chromosome of the ancient Paleo-Eskimo man was assigned to paragroup Q1a*, but the NWT01 locus was not sequenced. Assignment of the Paleo-Eskimo Y chromosome to Q1a6 does not conflict with these data or the TMRCA of Q1a6."*

[63] Reich *et al.* published a *corrigendum* in *Nature*, dated November 8, 2012, saying: *"At the time of publication of this Letter, the authors were unaware of a manuscript arriving at* **broadly similar**

7. Conclusion of Part I

What emerges from recent genetic research is a consensus on the idea that the prehistoric peopling of the Americas was the result of several migration waves, among which two bear a clear genetic signature: a major Paleo-Indian wave following a Pacific coastal route, generally dated back to ca. 16,000 BP, and a Paleo-Eskimo-Aleut wave in the northern extremes ca. 5,000 BP,[64] to which one should add a more recent Thule Eskimo migration ca. 1,000 AD. There is no consensus, however, concerning the contours of other gene flows, likely involving several populations and language groups, notably Na-Dené and Algonquian, but there is little doubt that more extensive studies of these populations, as well as of populations of the northern Pacific coast, will reinforce or refute a number of pending hypotheses.

Thus, on the basis of current knowledge, it appears that two components of Greenberg's tripartition hypothesis have been given support: a major, initial Amerindian migration wave possibly corresponding, on linguistic grounds, to the Amerind macrophylum, and one Paleo-Eskimo wave possibly corresponding to the Eskimoan linguistic family. The initial genetic pool delineates a unique, homogeneous founding population whose size, at the start of the migration process, has been estimated between 70 to a few hundred individuals (Hey 2005, Fagundes *et al.* 2007, 2008), a range hardly compatible with several languages. It is thus consistent to transpose these results into the linguistic domain by adopting the most parsimonious hypothesis, *i.e.* to equate the initial colonizing group with a single language, which one may call Amerind. This assumption provides a genetic and linguistic explanation frame to the distribution of *KOKO* kinship terms across the Americas, whereby *KOKO* terms were present in the kinship terminology of the initial Paleo-Indian group, thus confirming the views expressed in Greenberg, Turner & Zegura (1986), Greenberg (1987), Ruhlen (1994a, 1994b) and Greenberg & Ruhlen (2007) about the initial peopling of the Americas and in particular about the Amerind linguistic phylum.

Reconstructing the phonetic form of the Proto-Amerind *KOKO* kinship terms using the comparative method is still out of reach, but the number of existing or reconstructed reduplicated forms across the Americas, involving the velar consonant *k* and vowels *a* and *o* ~ *u*, strongly supports the original existence of the phonetic shapes *koko* ~ *kuku* ~ *kaka.*

The statistics of table 14, calculated from the data reported in table 13, unequivocally indicate that these forms were referring to a limited set of kinship relations which altogether display a clear parallelism between the feminine and the masculine sides. Although MB and EF appear as the preeminent relationships referred to by *koko* terms, at a continental scale, the GM and FZ relationships certainly do not represent random significations, nor do the Sib relationships. The task of the second part of the paper (Matthey de l'Etang and Bancel, to appear) will be precisely to unveil the semantics of *koko* kinship terms, and ultimately the nature of the kinship terminology of the first Amerinds.

---

conclusions *based on allotype analysis by Williams* et al., *which appeared in the* American Journal of Physical Anthropology *66, 1985."*

[64] Gilbert *et al.* (2008) and Achilli *et al.* (2013) disagree on the time frame regarding the introduction of the Beringian mtDNA haplogroup A2a into Northern Alaska, Northern Canada and Greenland: during the Thule expansion-wave ca. 1,000 AD according to Gilbert *et al.*, or during the Paleo-Eskimo wave ca. 5,000 BP according to Achilli *et al.* By contrast, both research teams agree that the mtDNA haplogroup D3 is probably associated with the Thule expansion.

REFERENCES

NB: For references with more than seven authors, only the first seven names are given, followed by the mention "*et al.*"

ADAM L. 1897. *Matériaux pour servir à l'établissement d'une grammaire comparée des dialectes de la famille Kariri.* Paris: Maisonneuve.

ADELAAR W.F.H. & P.C. MUYSKEN. 2004. *The languages of the Andes*. Cambridge (UK): Cambridge University Press.

ADELAAR W.F.H. 2012. Languages of the Middle Andes in areal-typological perspective. Emphasis on Quechuan and Aymaran. In CAMPBELL & GRONDONA, 575-624.

ACHILLI A., U.A. PEREGO, C.M. BRAVI, M.D. COBLE, Q-P. KONG, S.R. WOODWARD, A. SALAS *et al.* 2008. The phylogeny of the four pan-American mtDNA haplogroups. Implications for evolutionary and disease studies. *PLoS ONE* **3**-3: e1764. doi:10.1371/journal.pone.0001764.

ACHILLI A., U.A. PEREGO, H. LANCIONI, A. OLIVIERI, F. GANDINI, B. HOOSHIAR KASHANI, V. BATTAGLIA *et al.* 2013. Reconciling migration models to the Americas with the variation of North American Native mitogenomes. *PNAS* **110**-35, 14308-14313.

AIKHENVALD A.Y. 1999a. The Arawak language family. In DIXON & AIKHENVALD (1999b), 65-105.

AIKHENVALD A.Y. 1999b. Areal diffusion and language contact in the Içana-Vaupés basin, north-west Amazonia. In DIXON & AIKHENVALD, 385-416.

AIKHENVALD A.Y. 2002. *Language contact in Amazonia.* Oxford: Oxford University Press.

AMARANTE RIBEIRO L.A. 2006. *Uma proposta de classificação interna das línguas da família Pano.* www.unucseh.ueg.br/gicli/Lincoln/ArtigoInvestigacoes.pdf.

ANONBY S. 2012. A historical-comparative look at four Mondé languages. SIL: *Electronic Working Paper* **006**. www-01.sil.org/silewp/2012/silewp2012-006.pdf.

ANONBY S. & D.J. HOLBROOK. 2012. A report and a comparative-historical look at the Cinta Larga, Suruí, Gavião, and Zoró languages. *Working Papers of the Linguistic Circle of the University of Victoria* **23**, 14-31. http://journals.uvic.ca/index.php/WPLC/article/view/11908 (accessed dec. 2013).

ANS A-M. D' 1973. Reclasificación de las lenguas pano y datos glotocronológicos para la etnohistoria de la Amazonía peruana. *Revista del Museo Nacional* **39**, 349-369.

ANTUNES DE ARAUJO G. 2004. *A Grammar of Sabanê, a Nambikwaran language.* PhD diss. Utrecht: Landelijke Onderzoekschool Taalwetenschap.

AOKI. H. 1966. Nez Perce and Proto-Sahaptian kinship terms. *International Journal of American Linguistics*. **32**-4, 357-368.

ARRIAGA P.J. DE. 1621. *Extirpación de la idolatria del Piru.* Lima: Geronymo de Contreras.

ASHER R.E & C. MOSELEY (Eds.) 2007. *Atlas of the World's Languages*. London: Routledge.

BAILLIET G., M.MUZZIO, V. RAMALLO, L.S. JURADO MEDINA, E.L. ALFARO, J.E. DIPIERRI, & C.M BRAVI. 2012. Pre-Columbian male ancestors for the American continent. Molecular Y-chromosome insight. In M. CALISTAN (Ed.), *Genetic Diversity in Microorganisms*, InTech, 291-308.

BANCEL P.J. & A. MATTHEY DE L'ETANG. 2002. Tracing the ancestral kinship system. The global etymon *kaka*. Part I: A linguistic study. *Mother Tongue* **7**, 209-243.

BANCEL P.J. & A. MATTHEY DE L'ETANG. 2005. Kin tongue. A study of kin nursery terms in relation to language acquisition, with a historical and evolutionary perspective. *Mother Tongue* **9**, 171-190.

BANCEL P.J. & A. MATTHEY DE L'ETANG. 2013. Brave new words. In C. LEFEBVRE, B. COMRIE & H. COHEN (Eds.), *New Perspectives on the Origins of Language,* Amsterdam: John Benjamins, 333-377.

BANCEL P.J., A. MATTHEY DE L'ETANG, J. BENGTSON. 2011. Back to Proto-Sapiens (Part 2). The global kinship terms *papa*, *mama* and *kaka*. In D. JONES and B. MILICIC (Eds.), 38-43.

BARZANA A. DE. 1586. *Arte y vocabulario en la lengua general del Peru llamada Quichua, y en la lengua Española.* En los Reyes: Antonio Ricardo.

BELLEZA CASTRO N. 1995. *Vocabulario Jacaru-Castellano, Castellano-Jacaru (Aimara Tupino).* Cuzco: Centro de estudios regionales andinos «Bartolomé de las Casas».

BENGTSON J.D. (Ed.). 2008. *In hot pursuit of language in prehistory. Essays in the four fields of anthropology in honor of Harold Crane Fleming.* Amsterdam: John Benjamins.

BERTONIO L. 1603. *Arte y gramatica muy copiosa de la lengua Aymara.* Roma: Luis Zannetti.

BERTONIO L. 1612. *Vocabulario de la lengua Aymara.* Juli Pueblo, Chucuito: Francisco del Canto.

BISSO-MACHADO R., M.S. JOTA, V. RAMALLO, V.R. PAIXÃO-CÔRTES, D.R. LACERDA, F.M. SALZANO, S.L. BONATTO *et al.* 2011. Distribution of Y-chromosome Q lineages in Native Americans. *American Journal of Human Biology* **23**-4, 563-566.

BLENCH R.M. 2008. The problem of pan-African roots. In J.D. BENGTSON (Ed.), 189-209.

BODNER M., U.A. PEREGO, G. HUBER, L. FENDT, A.W. RÖCK, B. ZIMMERMANN, A. OLIVIERI *et al.* 2012. Rapid coastal spread of First Americans. Novel insight from South America's Southern Cone mitochondrial genomes. *Genome Research* **22**, 811-820.

BOLTON R. & E. MAYER. 1977. *Andean kinship and marriage.* Washington, DC: Amer. Anthropological Association.

BORTOLINI M.C, F.M. SALZANO, M.G. THOMAS, S. STUART, S.P.K. NASANEN, C.H.D. BAU, M.H. HUTZ *et al.* 2003. Y-chromosome evidence for differing ancient demographic histories in the Americas. *American Journal of Human Genetics* **75**, 524-39.

BRETON R. 1999 [1665]. *Dictionnaire caraïbe-français.* M. BESADA (Ed.). Paris: Karthala - IRD.

BRETON R. 1666. *Dictionnaire français-caraïbe.* Auxerre: G. Bouquet.

BRUNELLI G. 1987. Migrations, guerres et identité: faits ethno-historiques zoró. *Anthropologie et sociétés* **11**-3, 149-172.

BUENAVENTURA E. 1993. Observaciones preliminares acerca del idioma Macaguán. Apuntes culturales, fonología, apuntes gramaticales, vocabulario Macaguán-Español. Santafé de Bogotá (Colombia): Instituto Lingüístico de Verano.

CALLAGHAN C.A. 1997. Evidence for Yok-Utian. *International Journal of American Linguistics* **63**-1, 18-64.

CALLAGHAN C.A. 2001. More evidence for Yok-Utian. A reanalysis of the Dixon and Kroeber sets. *International Journal of American Linguistics* **67**-3, 313–345.

CAMPBELL L. 1997. *American Indian languages. The historical linguistics of Native America.* Oxford: Oxford U. Press.

CAMPBELL L. 2012. Classification of the indigenous languages of South America. In CAMPBELL & GRONDONA (2012), 59-166.

CAMPBELL L. & V. GRONDONA (Eds.) 2012. *The indigenous languages of South America. A comprehensive guide.* Berlin/Boston: De Gruyter Mouton.

CAMPBELL L. & D. OLTROGGE 1980. Proto-Tol (Jicaque). *International Journal of American Linguistics* **46**-3, 205-223.

CERRÓN-PALOMINO R. 1990. Reconsideración del llamado "quechua costeño." *Revista Andina* **8**-2, 335-409.

CHATTERS J.C, J. KENNETT, Y. ASMEROM, B.M. KEMP, V. POLYAK, A.N. BLANK *et al.* 2014. Late Pleistocene Human Skeleton and mtDNA Link Paleoamericans and Modern Native Americans. *Science* **344**, 750-4.

CHRISTIAN D.R. & E. MATTESON. 1972. Proto-Guahiban. In E. MATTESON (Ed.), 150-59.

COLLINS J.L. 1981. *Kinship and seasonal migration among the Aymara ou Southern Peru. Human adaptation to energy scarcity.* PhD diss. University of Florida.

CONSTENLA UMAÑA A. 1981. *Comparative Chibchan phonology.* PhD diss. University of Pennsylvania.

CONSTENLA UMAÑA A. 2002. Acerca de la relación genealógica de las lenguas Lencas y de las lenguas Misumalpas. *Filología y Lingüística* **28**-1, 189-205.

CONSTENLA UMAÑA A. 2009. Chibchan. In K. BROWN & S. OGILVIE (Eds.). *Concise Encyclopedia of Languages of the World.* Oxford, UK: Elsevier, 208-210.

CONSTENLA UMAÑA A. 2012. Chibchan languages. In CAMPBELL & GRONDONA (2012), 391-440.

COSTA NETO E.M. 2007. O Caranguejo-de-agua doce, *Trichodactylus fluviatilis* (Latreille 1828) (Crustacea, Decapoda, Trychodactylidae) na concepcão dos moradores do povoado de Pedra Branca, Bahia, Brasil. *Biotemas* **20**-1, 59-68.

CRAWFORD M.H., R.C. RUBICZ, & M. ZLOJUTRO. 2010. Origins of Aleuts and the genetic structure of populations of the archipelago: Molecular and archeological perspectives. *Human Biology* **82**-5/6, 695-717.

CREQUI-MONTFORT G. DE & P. RIVET. 1921. La famille linguistique takana. *Journal de la Société des américanistes de Paris* **13**-2, 281-301.

CREQUI-MONTFORT G. DE & P. RIVET 1922. La famille linguistique takana (suite). *Journal de la Société des américanistes de Paris* **14**-1, 141-182.

DARNELL R. & D. HYMES. 1986. Edward Sapir's six unit classification of American Indian languages. The search for time perspective. In T. BYNON & F.R. PALMER, *Studies in the history of Western linguistics in honour of R.H. Robins*. Cambridge, Eng.: Cambridge University Press, 202-244.

DENNY J.P. 1991. The Algonquian migration from Plateau to Midwest. Linguistics and archaeology. In W. COWAN (Ed.), *Papers of the Twenty-Second Algonquian Conference*. Ottawa: Carleton University Press, 103-124.

DIXON R.M.W. & A.Y. AIKHENVALD. 1999a. Introduction. In DIXON & AIKHENVALD (1999b), 1-22.

DIXON R.M.W. & A.Y. AIKHENVALD (Eds.) 1999b. *The Amazonian languages.* Cambridge (Eng.): Cambridge University Press.

DIXON R.M.W. & A.R. VOGEL. 2004a. *The Jarawara language of Southern Amazonia.* Oxford-New York: Oxford University Press.

DIXON R.M.W. 2004b. Proto-Arawa phonology. *Anthropological Linguistics* **46**-1, 1-83.

DULIK M.C., A.C. OWINGS, J.B. GAIESKI, M.G. VILAR, A. ANDRE, C. LENNIE, M.A. MACKENZIE *et al.* 2012. Y-chromosome analysis reveals genetic divergence and new founding native lineages in Athapaskan- and-Eskimoan-speaking populations. *PNAS* **109**-22, www.pnas.org/cgi/doi/10.1073/pnas.1118760109, 8471-8476.

DUVIOLS P. 1974. Une petite chronique retrouvée. Errores, ritos, supersticiones y ceremonias de los yndios de la provincia de Chinchaycocha y otras del Piru (1603). *Journal de la Société des américanistes de Paris* **63**, 275-297.

EBERHARD D.M. 2009. *Mamaindê grammar. A northern Nambikwara language and its cultural context,* Vols. 1 & 2. PhD diss. Utrecht: Landelijke Onderzoekschool Taalwetenschap.

ELIOT. J. 1663. *The holy Bible. Containing the Old Testament and the New. Translated into the Indian language.* Cambridge, MA: Green & Johnson.

ERIKSON P. 1986. Altérité, tatouage et anthropophagie chez les Pano. La belliqueuse quête du soi. *Journal de la Société des américanistes de Paris* **72**, 185-210.

ERIKSON P. 1993. Une nébuleuse compacte: le macro-ensemble pano. *L'Homme* **33**-126-128, 45-58.

FACUNDES S.S. da & A.P.B. BRANDÃO. 2011. Comparative Arawak linguistics. Notes on reconstruction, diffusion, and Amazonian prehistory. In A. HORNBORG & J.D. HILL (Eds.), *Ethnicity in ancient Amazonia.* Boulder: University Press of Colorado, 197-210.

FAGUNDES N.J.R., N. RAY, M. BEAUMONT, S. NEUENSCHWANDER, F.M. SALZANO, S.L. BONATTO & L. EXCOFFIER. 2007. Statistical evaluation of alternative models of human evolution. *PNAS* **104**, 17614-17619.

FAGUNDES N.J.R., R. KANITZ, R. ECKERT, A.C.S. VALLS, M.R. BOGO, F.M. SALZANO, D.G. SMITH *et al.* 2008. Mitochondrial population genomics supports a single pre-Clovis origin with a coastal route for the peopling of the Americas. *American Journal of Human Genetics* **82**, 1-10.

FORSTER P., R. HARDING, A. TORRONI, H-J. BANDELT. Origin and evolution of Native American mtDNA variation. A reappraisal. *American Journal of Human Genetics* **59**, 935-45.

FORTESCUE M., S. JACOBSON, & L. KAPLAN. 1994. *Comparative Eskimo dictionary with Aleut cognates.* Fairbanks: University of Alaska.

FLECK D.W. 2013. Panoan languages and linguistics. *American Museum of Natural History Anthropological Papers* **99**.

FOSTER M.K. 1996. Language and the culture history of North America. In W.C. STURTEVANT (Ed.), *Handbook of North American Indians.* Washington, D.C.: U.S. Government Printing Office.

FOWLER JR. W.R. 1989. *The cultural evolution of ancient Nahua civilizations. The Pipil-Nicarao of Central America.* Norman: University of Oklahoma Press.

GARCIA B. 1760. *Manual por administrar los sanctos sacramentos de penitencia eucaristía, extrema unción y matrimonio: dar gracia después de comulgar, y ayudar a bien morir.* México: Herederos de Doña Maria Rivera.

GILBERT M.T.P., T. KIVISILD, B. GRØNNOW, P.K. ANDERSEN, E. METSPALU, M. REIDLA, E. TAMM *et al.* 2008. Paleo-Eskimo mtDNA genome reveals matrilineal discontinuity in Greenland. *Science* **320**, 1787-1789.

GILBERTI M. 1901 [1559]. *Diccionario de la lengua tarascan ó de Michoacán.* A. PEÑAFIEL (Ed.). México: Oficina Impresora de Estampillas.

GILDEA S., B.J. HOFF, S. MEIRA. 2010. The story of **ô* in the Cariban family. In A.L. BEREZ, J. MULDER & D. ROSENBLUM (Eds.), *Fieldwork and linguistic analysis in indigenous languages of the Americas.* Honolulu: University of Hawai'i Press, 91-123.

GILDEA S. 2012. Linguistic studies in the Cariban family. In CAMPBELL & GRONDONA (2012), 441-494.

GIRARD V.J. 1971. *Proto-Takanan phonology.* Berkeley: University of California Press.

GODDARD I. 1994. *A new look for Algonquian.* Comparative Linguistics Workshop, University of Pittsburgh, April 9.

GOEJE C.H. DE. 1909. *Etudes linguistiques caraïbes.* Amsterdam: Johannes Müller.

GOEJE C.H. DE. 1946. *Etudes linguistiques caraïbes*, tome II. Amsterdam: Noord-Hollandsche Uitgevers.

GOLLA V. 2007. Linguistic Prehistory. In T.L. JONES and K.A. KLAR (Eds.), *California prehistory. Colonization, culture, and complexity.* New York: Altamira, 71-82.

GOLLA V., I. GODDARD, L. CAMPELL, M. MITHUN, M. MIXCO. 2007. North America. In ASHER & MOSELEY, 7-41.

GOODYEAR A.C. 2005. Evidence for pre-Clovis sites in the eastern United States. In R. BONNICHSEN, B.T. LEPPER, D. STANFORD & M.R. WATERS (Eds.), *Paleoamerican origins. Beyond Clovis.* College Station, TX: Texas A&M University Press, 103-112.

GONÇÁLEZ HOLGUÍN D. 1607. *Gramatica y arte nueva de la lengua general de todo el Peru, llamada lengua Qquichua, o lengua del Inca.* Ciudad de los Reyes: Francisco del Canto.

GONÇÁLEZ HOLGUÍN D. 1608. *Vocabulario de la lengua general de todo el Peru llamada lengua Qquichua, o del Inca.* Ciudad de los Reyes: Francisco del Canto.

GRAVIER J. ca. 1700. *Illinois-French dictionary.* Ms. Hartford, CT: Trinity College, Watkinson Library.

GREENBERG J.H. 1987. *Language in the Americas.* Stanford, CA: Stanford University Press.

GREENBERG J.H. & M. RUHLEN. 2007. *An Amerind etymological dictionary.* Version 12, Sept. 2007. Stanford, CA: Stanford University. http://www.merrittruhlen.com/files/AED5.pdf.

GREENBERG J.H., C.G. TURNER & S.L. ZEGURA. 1986. The settlement of the Americas. A comparison of the linguistic, dental, and genetic evidence. *Current Anthropology* **27**-5, 477-497.

GUTHRIE M. 1967-1971. *Comparative Bantu*, 4 vols. Farnborough, Hants: Gregg International.

HAGE P., B. MILICIC, M. MIXCO & J.P. NICHOLS. 2004. The Proto-Numic kinship system. *Journal of Anthropological Research* **60**-3, 359-377.

HASPELMATH M. 2004. How hopeless is genealogical linguistics, and how advanced is areal linguistics? (Review of A.Y. Aikhenvald & R.M.W. Dixon (Eds.), 2001. *Areal Diffusion and Genetic Inheritance.* Oxford: Oxford University Press), *Studies in Language* **28**-1, 209-223.

HARDMAN DE BAUTISTA M.J. 1969. Kawki: Chavin (Nicanor B. Hipolito Perez, questionnaire, Alipia Mateo Ascensio). University of Florida Digital Collection, http://ufdc.ufl.edu/UF00086721/00001.

HARDMAN DE BAUTISTA M.J. 1982. The mutual influence of Spanish and the Andean languages. *Word* **33**-1/2, 143-157.

HARDMAN M.J. 1983. Jaqaru short vowels. *International Journal of American Linguistics* **49**-2, 203-207.

HECKENBERGER M.J. 2002. Rethinking the Arawakan diaspora. Hierarchy, regionality, and the Amazonian formative. In HILL & SANTOS-GRANERO (2002), 100-122.

HEGGARTY P. 2008. Linguistics for archeologists. A case study in the Andes. *Cambridge Archeological Journal* **18**-1, 35-56.

HELGASON A., G. PÁLSSON, H.S. PEDERSEN, E ANGULALIK, E.D. GUNNARSDÓTTIR, B. YNGVADÓTTIR, K. STEFÁNSSON. 2006. MtDNA variation in Inuit populations of Greenland and Canada. Migration history and population structure. *American Journal of Physical Anthropology* **130**-1, 123-134.

HEY J. 2005. On the number of New World founders. A population genetic portrait of the peopling of the Americas. *PLoS Biology*, www.plosbiology.org0965, Vol. **3**-6: e193, 0965-0975.

HILL J.H. 2010. New evidence for a Mesoamerican homeland for Proto-Uto-Aztecan. *PNAS* **107**-11. www.pnas.org/cgi/doi/10.1073/pnas.0914473107.

HILL J.H. & F. SANTOS-GRANERO. 2002. *Comparative Arawakan history. Rethinking language family and culture area in Amazonia.* Urbana-Champaign: University of Illinois Press.

HOPKINS N.A. 1984. Otomanguean linguistic prehistory. In JOSSERAND J.K., M. WINTER & N.A. HOPKINS (Eds.), *Essays in Otomanguean culture history*. Nashville: Vanderbilt University, 25-64.

HOOPES J.W. & O.M. FONSECA Z. 2003. Goldwork and Chibchan identity. Endogenous changes and diffuse unity in the Isthmo-Colombian area. In J. QUILTER & J.W. HOOPES (Eds.), *Gold and power in ancient Costa Rica, Panama, and Colombia.* Washington, DC: Dumbarton Oaks Research Library, 49-89.

JOLKESKY DE VALERY M.P. 2010. *Reconstrução fonológica e lexical do proto-jê meridional.* MA diss. Universidad estadual de Campinas.

JONES D. & B. MILICIC (Eds.). 2011. *Kinship, language, and prehistory. Per Hage and the renaissance of kinship studies.* Salt Lake City: University of Utah Press.

KAUFMAN T. 1976. Archaeological and linguistic correlations in Mayaland and associated areas of Meso-America. *World Archaeology* **8**-1 "Archaeology and Linguistics", 101-118.

KAUFMAN T. 1994a [2nd edition 2007a]. The Native languages of South America. In ASHER & MOSELEY (2007), 61-78.

KAUFMAN T. 1994b [2nd edition 2007b]. The Native languages of Meso-America. In ASHER & MOSELEY (2007), 47-54.

KAUFMAN T. & J. JUSTESON. 2003. *A preliminary Mayan etymological dictionary.* www.famsi.org/reports/01051/pmed.pdf.

KAUFMAN T. & J. JUSTESON. 2009. Historical linguistics and pre-Columbian Mesoamerica. *Ancient Mesoamerica* **20**-2, 221-231.

KEELS J. 1985. Guayabero. Phonology and morphophonemics. In R. BREND (Ed.), *From phonology to discourse. Studies in six Colombian languages.* Amerindian Series **9**. Dallas: Summer Institute of Linguistics, 57-87.

KEELS J. 1986. *Proto-Guahiban. A comparative reconstruction of Cuiva, Guahibo and Guayabero,* ms. Bogotá: Instituto Lingüistico de Verano.

KEMP B.M., R.S. MALHI, J. MCDONOUGH, D.A. BOLNICK, J.A. ESHLEMAN, O. RICKARDS, C. MARTINEZ-LABARGA *et al.* 2007. Genetic analysis of early Holocene skeletal remains from Alaska and its implications for the settlement of the Americas. *American Journal of Physical Anthropology* **132**, 605-621.

KENSINGER K.M. 1995. *How real people ought to live. The Cashinahua of Eastern Peru.* Prospect Heights, IL: Waveland Press.

KEY M.R. 1968. *Comparative Tacanan phonology with Cavineña phonology and notes on Pano-Tacanan relationship.* The Hague: Mouton.

KEY M.R., R.M. TUGWELL, M. WESSELS 1992. Araona correspondences in Tacanan. *International Journal of American Linguistics* **58**-1, 96-117.

KINKADE M.D. 1991. Prehistory of the native languages of the Northwest Coast. In E.A.P. CROWNHART-VAUGHAN (Ed.), *Proceedings of the Great Ocean Conference,* Vol. 1: *the North Pacific to 1600.* Portland: Oregon Historical Society Press, 137-158.

KINKADE M.D. 1992. Kinship terminology in Upper Chehalis in a historical framework. *Anthropological Linguistics* **34**-1/4, 84-103.

KOCH-GRÜNBERG T. 1911. *Aruak-Sprachen Nordwestbrasiliens und der angrenzenden Gebiete.* Vienna: Anthropologische Gesellschaft.

KOONTZ-GARBODEN A. & I. FRANCEZ 2009. *Possessed properties in Ulwa.* http://personalpages.manchester.ac.uk/staff/andrewkg/koontz-garboden-francez-poss-pc-nals-preprint.pdf.

KRAUSS M.E. 1973. Na-Dene. In T.A. SEBEOK (Ed.), *Linguistics in North America*, Current Trends in Linguistics **10**. The Hague: Mouton, 903-978.

KROEKER M.H. 2003. *Gramática descritiva da língua nambikuara.* Cuiabá: Sociedade internacional de lingüística. http://www-01.sil.org/americas/brasil/publcns/dictgram/NBGram.pdf.

KUMAR S., C. BELLIS, M. ZLOJUTRO, P.E. MELTON, J. BLANGERO & J.E. CURRAN. 2011. Large-scale mitochondrial sequencing in Mexican Americans suggests a reappraisal of Native American origins. *BMC Evolutionary Biology* **11**-293, 1-17.

LANDAR H. 1976-1977. South and Central American Indian languages. In T.A. SEBEOK (Ed.), *Native Languages of the Americas*, Vol. 2. New York: Plenum Press, 401-527.

LELL J.T., R.I. SUKERNIK, Y.B. STARIKOVSKAYA, B. SU, L. JIN, T.G. SCHURR, P.A. UNDERHILL & D.C. WALLACE. 2002. The dual origin and Siberian affinities of Native American Y-chromosomes. *American Journal of Human Genetics* **70**, 192-206.

MACEDO REIS DE ALMEDA BRITO A. 2005. Reconstrução lexical e fonológica do Proto-Mondé. 2ª parte (comunicaçaõ oral: resumo 56). *O saber como estratégia de vida e o desenvolvimento da Amazônia.* 13° Seminário de Iniciação Científica. PIBIC, 28 a 30 de junho de 2005. Belém: Museu Paraense Emílio Goeldi.

MACEDO REIS DE ALMEDA BRITO A. 2006. Reconstrução lexical e fonológica do Proto-Mondé. 2ª parte (comunicaçaõ oral: resumo 27*). Ciência e tecnologia no contexto amazônico.* 14° Seminário de Iniciação Científica. PIBIC, 01 a 03 de agosto de 2006. Belém: Museu Paraense Emílio Goeldi.

MALYARCHUK B., M. DERENKO, G. DENISOVA, A. MAKSIMOV, M. WOZNIAK, T. GRZYBOWSKI, I. DAMBUEVA *et al.* 2011. Ancient links between Siberian and Native Americans revealed by subtyping the Y-chromosome haplogroup Q1a. *Journal of Human Genetics* **56**, 583-588.

MAMIANI L.V. 1698. *Catecismo da doutrina christãa na lingua brasilica da nação kiriri.* Lisbon: Miguel Deslandes.

MAMIANI L.V. 1699. *Arte de grammatica da lingua brasilica da naçam kiriri.* Lisbon: Miguel Deslandes.

MANRIQUEZ G., M. MORAGA, C. SANTORO, E. ASPILLAGA, B.T. ARRIAZA & F. ROTHHAMMER. 2011. Morphometric and mtDNA analyses of archaic skeletal remains from southwestern South America. Análisis morfométrico y de ADNmt de restos esqueletales arcaicos del Suroeste de Sudamérica. *Chungara. Revista de Antropología Chilena* **43**-2, 283-292.

MARTIUS C.F.P VON. 1867*a*. *Beiträge zur Ethnographie und Sprachenkunde Amerika's zumal Brasiliens,* Bd I. *Zur Ethnographie Amerika's zumal Brasiliens.* Leipzig: Friedrich Fleischer.

MARTIUS C.F.P VON. 1867*b*. *Beiträge zur Ethnographie und Sprachenkunde Amerika's zumal Brasiliens,* Bd II. *Wörtersammlung brasilianischer Sprachen. Glossaria linguarum brasiliensium. Glossarios de diversas lingoas e dialectos, que fallao os Indios no imperio do Brazil.* Leipzig: Friedrich Fleischer.

MATTESON E. 1972a. Proto-Arawakan. In MATTESON (1972), 160-242.

MATTESON E. 1972b. *Comparative studies in Amerindian languages.* The Hague: Mouton.

MATTHEWS G.H. 1959. Proto-Siouan kinship terminology. *American Anthropologist* **61**-2, 252-278.

MATTHEY DE L'ETANG A. & P.J. BANCEL. 2002. Tracing the ancestral kinship system: The global etymon *kaka*. Part II: An anthropological study. *Mother Tongue* 7, 245-258.

MATTHEY DE L'ETANG A. & P.J. BANCEL. 2008. The age of *mama* and *papa*. In BENGTSON (2008), 417-438.

MATTHEY DE L'ETANG A. & P.J. BANCEL. To appear. The kinship term *kuku ~ koko ~ kaka* in the American Indigenous languages, the Amerind hypothesis, and the Dravidian kinship system. Part II: Anthropological Study.

MATTHEY DE L'ETANG A. & P.J. BANCEL. In preparation. *Borrowing of kin terms in indigenous South American languages.*

MATTHEY DE L'ETANG A., P.J. BANCEL & M. RUHLEN. 2011. Back to Proto-Sapiens (Part 1). The inherited kinship terms *papa, mama* and *kaka*. In JONES & MILICIC (2011), 29-37.

MEEUSSEN A. E. 1969. *Bantu lexical reconstructions*. http://www.linguistics.berkeley.edu/CBOLD/Docs/Meeussen.html.

MEIRA S. 1998. *A reconstruction of Proto-Taranoan. Phonology and inflectional morphology*. MA diss. Houston, TX: Rice University.

MERRIFIELD W.R. 1981. *Proto-Otomanguean kinship*. Dallas, TX: Summer Institute of Linguistics.

MERRILL W.L., R.J. HARD, J.B. MABRY, G.J. FRITZ, K.R. ADAMS, J.R. RONEY & A.C. MACWILLIAMS. 2009. The diffusion of maize to the southwestern United States and its impact. *PNAS* **106**-50, 21019-21026. http://www.pnas.org/content/106/50/21019.full.pdf+html.

MIHAS E. 2010. *Essentials of Ashéninka Perené grammar*. PhD diss. Milwaukee: University of Wisconsin.

MILICIC B. 2011. Is there a kinship module? Evidence from children's acquisition of kinship terms in Pitumarca, Peru. In JONES & MILICIC (2011), 212-221.

MILLER W.R. 1967. *Uto-Aztecan cognate sets*. Berkeley: University of California Press.

MOORE D. 2005. Classifição interna da família lingüística Mondé. *Estudos Lingüísticos* **34**, 515-520.

MORGAN L.R. 1980. *Kootenay-Salishan linguistic comparison. A preliminary study*. M.A. diss. University of British Columbia.

MUDRAK Oleg A. 2005b. *Eskimo etymology*. Online etymological database, Tower of Babel Project. http://starling.rinet.ru/cgibin/main.cgi?flags=eygtnnl.

MYERS T.P. 1974. Spanish Contacts and Social Change on the Ucayali River, Peru. *Ethnohistory* **21**-2, 135-157.

NANTES B. DE. 1709. *Katecismo Indico da Lingua Kariris*. Lisbon: Valentim da Costa Deslandes.

NANTES M. DE. n.d. *Relation succinte et sincère de la mission du père Martin de Nantes, prédicateur capucin, missionnaire apostolique dans le Brezil parmy les indiens appellés Cariris*. Quimper: Jean Perier.

NIKOLAYEV S.L. 2007. *Indo-European Etymology*. Online etymological database, Tower of Babel Project. http://starling.rinet.ru/cgi-bin/main.cgi?flags=eygtnnl.

NOBLE G.K. 1965. *Proto-Arawakan and its descendants*. Bloomington: Indiana University Press.

PAYNE D.L. 1981. *The phonology and morphology of Axininca Campa*. Arlington, TX: University of Texas.

PAYNE D.L. 1987. Some morphological elements of Maipuran Arawakan. Agreement affixes and the genitive construction. *Language Sciences* **9**-1, 57-75.

PAYNE D.L. 1991. A classification of Maipuran (Arawakan) languages based on shared lexical retentions. In D.C. DERBYSHIRE & G.K. PULLUM (Eds.), *Handbook of Amazonian languages* **3**. Berlin: Mouton de Gruyter, 355-499.

PEREGO U.A., A. ACHILLI, N. ANGERHOFER, M. ACCETTURO, M. PALA, A. OLIVIERI, B. HOOSHIAR KASHANI *et al.* 2009. Distinctive Paleo-Indian migration routes from Beringia marked by two rare mtDNA haplogroups. *Current Biology* **19**, 1-8.

PYLE R.C 1981. Aymara kinship, real and spiritual. In M.J. HARDMAN (Ed.). *The Aymara language in its social and cultural context*. Gainesville: University Presses of Florida, 80-89.

QUINTANA A. DE. 1733. *Confessonario en lengua Mixe*. En la Puebla: Viuda de Miguel de Ortega.

RADIN P. 1931. *Mexican kinship terms*. University of California Publications in American Archeology and Ethnology **31**-1, 1-14.

RAFF J.A., M. RZHETSKAYA, L. ARMSTRONG & M.G. HAYES. 2011. *Increased genetic diversity observed among Inupiat populations from the North Alaskan slope*. Poster presented to the American Society of Human Genetics 61st Annual Meeting, Oct. 14. 2011, session "Evolutionary and Population Genetics".

RAGHAVAN M., P. SKOGLUND, K. E. GRAF, M. METSPALU, A. ALBRECHTSEN, I. MOLTKE, S. RASMUSSEN *et al.* 2014. Upper Palaeolithic Siberian genome reveals dual ancestry of Native Americans. *Nature* **505**, 87-91.

RASLES S. 1691 [1833]. *A Dictionary of the Abnaki Language, in North America*. Offprint from Memoirs of the American Academy of Arts and Sciences New Series, Vol. 1, 370-574.

RASMUSSEN M., Y. LI, S. LINDGREEN. J.S. PEDERSEN, A. ALBRECHTSEN, I. MOLKTE, M. METSPALU *et al.* 2010. Ancient human genome sequence of an extinct Palaeo-Eskimo. *Nature* **463**-11, 757-762.

REICH D., N. PATTERSON, D. CAMPBELL, A. TANDON, S. MAZIERES, N. RAY, M.V. PARRA *et al.* 2012. Reconstructing Native American population history. *Nature* **468**, 370-374.

RIVET P. 1948. La famille linguistique guahibo. *Journal de la Société des américanistes de Paris* **37**, 191-240.

RIVET P. 1949. Les langues de l'ancien diocèse de Trujillo. *Journal de la Société des américanistes de Paris* **38**, 1-51.

RODRIGUES A.D. 1986. *Línguas brasileiras: para o conhecimento das línguas indígenas*. São Paulo: Loyola.

RODRIGUES A.D. 1999. Macro-Jê. In DIXON & AIKHENVALD (1999b), 165-206.

ROEWER, M. NOTHNAGEL, L. GUSMÃO, V GOMES, M GONZÁLEZ, D. CORACH, A. SALA *et al.* 2013. Continent-wide decoupling of Y-chromosomal genic variation from language and geography in Native South Americans. *PloS Genetics* **9**-4, 1-16.

ROMNEY A.K. 1967. Kinship and family. In R. WAUCHOPE (Ed.). *Handbook of Middle American Indians*, Vol. 6. Austin: University of Texas Press, 207-237.

RUHLEN M. 1994a. *The origin of language. Tracing the evolution of the mother tongue*. New York: John Wiley.

RUHLEN M. 1994b. The linguistic origins of Native Americans. In M. RUHLEN (Ed.), *On the origin of languages. Studies in linguistic taxonomy*. Stanford: Stanford University Press, 207-241.

SANTO TOMAS D. DE. 1560a. *Lexicon, o vocabulario de la lengua general del Peru*. Valladolid: Francisco Fernandez de Cordoua.

SANTO TOMAS D DE. 1560b. *Grammatica, o arte de la lengua general de los Indios de los Reynos del Peru*. Valladolid: Francisco Fernandez de Cordoua.

SAPIR E. 1922. The Algonkin Affinity of Yurok and Wiyot Kinship Terms. *Journal de la Société des américanistes*. **14-15**, 36-74.

SCHEFFLER H.W. 1978. *Australian kin classification*. Cambridge, MA: Cambridge University Press.

SCHROEDER K.B., T.G. SCHURR, J.C. LONG, N.A. ROSENBERG, M.H. CRAWFORD, L.A. TARSKAIA, L.P. OSIPOVA *et al.* 2007. A private allele ubiquitous in the Americas. *Biology Letters* **3**, 218-223.

SCHROEDER K.B., M. JAKOBSSON, M.H. CRAWFORD, T.G. SCHURR, S.M. BOCA, D.F. CONRAD, R.Y. TITO *et al.* 2009. Haplotypic background of a private allele at high frequency in the Americas. *Molecular Biology and Evolution* **26**-5, 995-1016.

SCHULLER R.R. 1928. *Las lenguas indígenas de Centro América con especial referencia a los idiomas aborígenes de Costa Rica.* San José de Costa Rica: Imprenta Nacional.

SEKI L. 1999. The Upper Xingu as an incipient linguistic area. In DIXON & AIKHENVALD (1999b), 417-430.

SHELL O.A. 2008 (1965). *Las lenguas Pano y su reconstrucción.* Serie Lingüística Peruana, Publicación Numeró **12**. Yarinacocha, Perú: Instituto Lingüístico de Verano.

SHEPERD A. 2006. *Proto-Wintun*. Los Angeles: University of California Press.

SIEBERT JR. F.T. 1967. The original home of Proto-Algonquian people. *Contributions to Anthropology: Linguistics I (Algonquian)*. Anthropological Series **78**. Ottawa: National Museum of Canada, Bulletin **214**, 13-47.

SILVY A. ca 1678. *Dictionnaire montagnais.* Ms. Ottawa: Canadian Government Archives. http://www.collectionscanada.gc.ca/008/001/008001-119.01-f.php?&document_id_nbr=101&ts_nbr=9&brws=1&&PHPSESSID=sra1gfoob09926b5s3u4go7u53 .

SMITH D.G., R.S. MALHI, J.A. ESHLEMAN, F.A. KAESTLE & B.M. KEMP. 2005. Mitochondrial DNA haplogroups of Paleoamericans in North America. In R. BONNICHSEN, B.T. LEPPER, D. STANFORD & M.R. WATERS (Eds.), *Paleoamerican origins. Beyond Clovis*. College Station, TX: Texas A&M University Press, 243-254.

SORENSEN Jr. A.P. 1967. Multilingualism in the Northwest Amazon. *American Anthropologist* **69**-6, 670-684.

SPIX J.B. VON & C.F.P .VON MARTIUS 1828. *Reise in Brasilien auf Befehl Sr. Majestät Maximilian Joseph I., Königs von Baiern, in den Jahren 1817 bis 1820.*, 2. Theil. Munich: I.J. Lentner.

STARIKOVSKAYA E.B., R.I. SUKERNIK, O.A. DERBENEVA, N.V. VOLODKO. E. RUIZ-PESINI, A. TORRONI, M.D. BROWN *et al.* 2005. Mitochondrial DNA diversity in indigenous populations of the southern extent of Siberia, and the origins of Native American haplogroups. *Annals of Human Genetics* **69**, 67-89.

STAROSTIN S.A. 2005c. *Sino-Tibetan etymology.* Online etymological database, Tower of Babel Project. http://starling.rinet.ru/cgi-bin/main.cgi?flags=eygtnnl.

STENZEL K. 2005. Multilingualism in the northwest Amazon revisited. *Memorias del congreso de idiomas indígenas de Latinoamérica II,* held 27-29 Oct. 2005 at Austin. http://www.ailla.utexas.org/site/cilla2/Stenzel_CILLA2_vaupes.pdf.

SWADESH M. 1958. Some new glottochronological dates for American linguistic groups. *Proceedings of the International Congress of Americanists* **32**. Copenhagen: Munskgaard, 671-674.

TAMM E., T. KIVISILD, M. REIDLA, M. METSPALU, D. GLENN SMITH, C.J. MULLIGAN, C.M. BRAVI *et al.* 2007. Beringian standstill and spread of Native American founders. *PloS One* **9**: e829. 1-6. doi:10.1371/journal.pone.0000829.

TARAZONA-SANTOS E. & F.R. SANTOS. 2002. The peopling of the Americas: A second major migration? *American Journal of Human Genetics* **70**, 1377-1380.

TAYLOR A.R. 1963. Comparative Caddoan. *International Journal of American Linguistics* **29**-2, 113-131.

TELLES S. & L. WETZELS. 2006. Evidentiality and epistemic mood in Lakondê. In E. CARLIN & G. ROWICKA (Eds.), *What's in a verb? Dutch contributions to the study of verbal morphology in the languages of the Americas.* Utrecht: Landelijke Onderzoekschool Taalwetenschap, 235-52.

TELLES ARAUJO PEREIRA LIMA S. 2002. *The phonology and grammar of Latundê/Lakondê.* PhD diss. Amsterdam: Vrije Universiteit.

TORRONI A., T.G. SCHURR, C-C YANG, E.J.E. SZATHMARY, R.C. WILLIAMS, M.S. SCHANFIELD, G.A. TROUP *et al.* 1992. Native American mitochondrial DNA analysis indicates that the Amerind and the Nadene populations were founded by two independent migrations. *Genetics* **130**, 153-62.

TRASK L. 2004. Where do *mama/papa* words come from? *University of Sussex Working Papers in Linguistics and English Language.* https://www.sussex.ac.uk/webteam/gateway/file.php?name=where-do-mama2.pdf&site=1.

TRUMBULL J.H. 1903. *Natick dictionary.* Smithsonian Institution, *Bureau of American Ethnology Bulletin* **25**. Washington: Government Printing Office.

VILLALÓN M.E. 1991. A spatial model of lexical relationships among fourteen Cariban varieties. In M.R. KEY (Ed.). *Language change in South American Indian languages.* Philadelphia: University of Pennsylvania Press, 54-94.

VOEGELIN C.F., F.M. VOEGELIN & K.L. HALE. 1962. Typological and comparative grammar of Uto-Aztecan, vol. 1: Phonology. *Supplement to International Journal of American Linguistics* **28**-1.

VOLODKO N.V., E.B. STARIKOVSKAYA, I.O. MAZUNIN, N.P. ELTSOV, P.V. NAIDENKO, D.C. WALLACE & R.I. SUKERNIK. 2008. Mitochondrial genome diversity in Arctic Siberians, with particular reference to the evolutionary history of Beringia and Pleistocenic peopling of the Americas. *American Journal of Human Genetics* **82**, 1084-1100.

WALKER H. 2009. Transformations of Urarina kinship. *Journal of the Anthropological Society of Oxford Online* **1**-1, 52-69.

WANG S., C.M. LEWIS JR., M. JAKOBSSON, S. RAMACHANDRAN, N. RAY, G. BEDOYA, W. ROJAS, M.V. PARRA *et al.* 2007. Genetic variation and population structure in Native Americans. *PLoS Genetics* **3**-11: e185. doi:10.1371/journal.pgen.0030185, 2049-67.

WEBSTER S.S. 1977. Kinship and affinity in a native Quechua community. In BOLTON & MAYER (1977), 28-42.

WHEELER C.J. 1982. An enquiry into the Proto-Algonquian system of social classification and marriage. A possible system of symmetric prescriptive alliance in a Lake Forest archaic culture during the third millennium BC. *Journal of the Anthropological Society of Oxford* **13**, 165-74.

WHISTLER K.W. 1977. Wintun prehistory. An interpretation based on linguistic reconstruction of plant and animal nomenclature. *Proceedings of the third annual meeting of the Berkeley Linguistic Society, February 19-21*. Berkeley, CA: Berkeley Linguistic Society, 157–174.

WHISTLER K.W. 1980. *Proto-Wintun kin classification. A case study in reconstruction of a complex semantic system*. PhD diss. Berkeley, CA: University of California.

WICHMANN S. 1999. On the relationship between Mixe-Zoquean and Uto-Aztecan. In X J. KYLE (Ed.). *Kansas Working Papers in Linguistics.* **24**-2, *Studies in Native American Linguistics* **10**, 101-113.

WICHMANN S., A. MÜLLER & V. VELUPILLAI. 2010. Homelands of the world's language families. A quantitative approach. In S. WICHMANN & A.P. GRANT (Eds.), *Quantitative approaches to linguistic diversity. Commemorating the centenary of the birth of Morris Swadesh, Diachronica* special issue **27**-2, 246-276.

WICHMANN S. & C.H. BROWN. N.d. Syllable nuclei of Proto-Mayan disyllabic stems. http://email.eva.mpg.de/~wichmann/WichmannBrownDisyllables3.pdf.

WILLIAMS R.C., A.G. STEINBERG, H. GERSHOWITZ, P.H. BENNETT, W.C. KNOWLER, D.J. PETTITT, W. BUTLER *et al.* 1985. GM allotypes in Native Americans. Evidence for three distinct migrations across the Bering land bridge. *American Journal of Physical Anthropology* **66**, 1-19.

YAYA I. 2008a. The importance of initiatory ordeals. Kinship and politics in an Inca narrative. *Ethnohistory* **55**-1, 51-85.

YAYA I. 2008b. *The two faces of Incan history. Unravelling dual representations in oral tradition of Pre-Hispanic Cuzco*. PhD diss. Sydney: University of New South Wales.

ZAMBRANO BONILLA J. 1752. *Arte de lengua Totonaca conforme á el arte de Antonio Nebrija*. La Puebla: Viuda de Miguel de Ortega.

ZEGURA S.L., T.M. KARAFET, L.A. ZHIVOTOVSKY & M.F. HAMMER. 2004. High-resolution SNPs and microsatellite haplotypes point to a single, recent entry of Native American Y-chromosomes into the Americas. *Molecular Biology and Evolution* **21**-1, 164-175.

ZLOJUTRO M. 2006-2008. *Mitochondrial DNA and Y-chromosome variation of Eastern Aleut populations. Implications for the genetic structure and peopling of the Aleutian archipelago*. PhD diss. Lawrence: University of Kansas.

ZUIDEMA R.T. 1977. The Inca kinship system. A new theoretical view. In BOLTON & MAYER (1977), 240-281.

ZWARTJES O. 2011. *Portuguese missionary grammars in Asia, Africa and Brazil, 1550-1800*. Studies in the History of the Language Sciences **117**. Amsterdam: John Benjamins.

APPENDIX

List of *KOKO* terms in South American language families. The classification is that of Campbell (2012).

Arawakan (Maipuran): **-kuko* U (MB), EF, **-aku[ro]* A (FZ), EM [Payne 1991]; **South Arawak:** Terena *eungo*, *eúko* Un, EF [Oberg 1948a , Oberg 1949, Parker 1987 in Payne 1991]; Kinikinau *ēukɔ* (♂)U, *haʔa* F [De Souza 2007]; Bauré *kik* MB, EF, *-aki* EM [Baptista & Wallin 1964]; Mucoxeone *ni-kiko* U [Créqui-Montfort & Rivet 1922, Rivet 1948]; Ignaciano (Moxo) *kuko*, *ne-kuko*, *-ekuka* U, *-aca* [-aka] A, GMZ [Créqui-Montfort & Rivet 1922, Rivet 1948, Ott 1983]; Trinitario *-očko* GF [Gill's field materials n.d.]; **Paresí-Xingu:** Mehináku *kukú* (♀)MB, *aki* FZ, EM [Schuller 1910, Galvão 1953, Gregor 1977]; Yawalapití *kukúzu* (♀)MB, *niʔua* my MB (ego ♂), *aáki* FZ ad., *na'k̯iru* my A [Oberg 1953, Galvão 1953, Ortega Mujica 1992]; Kustenau *húauá* MB [Schuller 1910]; Saraveca *koko-re-ixi* U, *kokore* F [Créqui-Montfort & Rivet 1922, Rivet 1948]; Paressi *koko* ~ *kuku* ad. U, EF, *koko-re* ~ *kuku-re* my U, *ene-koke* U, *enakero*(*u*) A, *naika* WM, *xaxa* my eZ, *axexe* my eB [Schuller 1910, Créqui-Montfort & Rivet 1922, Rivet 1948, Roquette Pinto 1975, Rowan & Burgess 1979, Rowan 2001]; **South-Western Arawak:** Piro (Yine) *kuku* ~ *koko* ad. MB, FZH, EF, *nukoxiru* my MB, my FZH, my HF [Schuller 1910, Créqui-Montfort & Rivet 1922, Löffler & Baer 1974]; Maxinéri *kooko* MB [Santos Silva 2008]; Apuriñá *kükü* ~ *kįkį* ~ *keké* man, *ukõku-ru* U of, *nakįru*, *akuro* GM [Koch-Grünberg 1914-1919, Créqui-Montfort & Rivet 1922, Nimuendajú in Archives Rivet, Rivet 1948, Nimuendajú 1955, Facundes 2000]; Iñapari *najɨró* my GM [Parker 1995]; Maschco Piro *kokoa* U [Ruhlen 1994]; Sirineiri *kokoa* U [Rivet 1948]; Cushichineri *koko* U [Archives Rivet, Rivet 1948]; Cuniba *kuku* MB, EF [Schuller 1910, Créqui-Montfort & Rivet 1922, Rivet 1948]; Kanamaré *ghughu* MB [Créqui-Montfort & Rivet 1922, Rivet 1948]; **Campa:** Campa Axininca *kokoini* ad. (♀)MB, HF, (♀) FZH, *koŋki* ad. (♂)MB, WF, (♂)FZH *no-koŋkiri* my WF, *no-koŋkit^{h}oriri* my distant U [Payne 1981]; Campa Asháninca *kokoxi* (♀)MB, *koki* MB, EF, *pikonkiri* your MB, your EF, *irairo* FZ, WF, (♂)FBW [Kindberg 2008]; Campa Asháninca (Tambo dial.) *kóóko* my MB (ego ♀), my HF, *kokí* my MB (ego ♂), my WF, *áironci* my WM, my FZ (ego ♂) [Weiss 1975]; Ashéninca (rio Apurucayali) *koko*, *kokoini* ad. (♀)U, HF, (♀)FZH, *koŋki* ~ *koki* ad. (♂)U, WF, (♂)FZH, *no-koŋkiri* my U, my EF, my FZH, *nayiro* (my) FZ, (my) HM, (my) MBW [Payne 1980, 1982]; Ashéninca (rio Ucayali) *koko*, *kokoini* ad. (♀)U, HF, (♀)FZH, *koŋki ~koki* ad. (♂)U, WF, (♂)FZH, *nokoŋkiri* my U, my EF, FZH, *nayiro* (my) FZ, (my) HM, (my) MBW [Payne 1980]; Ashéninka Perené *koki* ad. (♂)MB, WF, *kooko* ad. (♀)MB, *nokonkiri* MB, EF, *airontsi* ad. FZ (ego ♂), WM, *nayiro* EM [Mihas 2010, 2012]; Caquinte *kooŋkiini* tio [Swift 1988]; Matsiguenga *koki* MB, EF, FZH, *pikónkiri* my MB, my EF, my FZH, *nogókine* my MB, my EF, *pagiro* FZ, EM, MBW [Casevitz 1977, Snell 2008 (1998)]; Nomatsiguenga *kokí* ad. (♂)MB, (♂)EF, *no-kongiri* my MB, my EF, my FZH, *nágiro* my FZ, my EM [Shaver 1996]; Nanti *igoŋkirite* his WF, *pagiro* ad. EM, *obagirote* her HM [David Michael 2008]; **Amuesha:** Amuesha *nego'* MB, FZH, EF [Santos-Granero 1991]; **Chamicuro:** Chamicuro *-kohka* my U, *-ahka* (♀)A, *lokoko* GF [Parker 1987, 1991]; **Rio Branco:** Wapishana *kukúi* GM [Archives Rivet, Rivet 1948, Nimuendajú 1955, Diniz 1968, Melville et al. 2007]; **Palikur:** Palikur *nukukrin* (my) MB, EF, *kūku* ad. MB, EF, *nakirun*, (my) FZ, EM, *akiá* ad. id. [Nimuendajú 1926, Rivet 1948, Arnaud 1968]; Marawá *oky* ~ *ukí* U, *uhí* A [Martius 1867b, Archives Rivet: 1920]; **Ta-Arawak:** Guajiro (Wayuu) *ta-ʔirɯ* my A [Wavrin 1932 in Landaburu 1998-2, Wilbert 1958, Goulet 1981, Ramirez 1996 in Landaburu 1998-2]; Paraujano (Añu) *táurï* EF, *táira* FZ [Wilbert 1983]; Lokono (Arawak) *dakuru*, *akiru* ~ *akürŭ* my (♀)EM [Goeje 1928, Kirchhoff 1931, Hickerson 1953]; **Caribbean:** Island Carib *nøkɛka-jɛm* my MB, my FZH, *nakre* ~ *takɛre* my SW, *aka-tobu* MB [Breton 1665, 1666]; Black Carib *nagúru* HM, *nágoro* SW [Solien 1960, Taylor 1965]; **North-Amazonian:** Resígaro *-khíigi* MB, *-oʔhoogí* GF [Allin 1979]; Yucuna (Matapi) *oku* ~ *u'küh* MB, EF, *nokúru* FZ, *tx̱utx̱ú* GP [Koch-Grünberg 1911, Ortiz-Gomez 1983, Schauer & Schauer 1978, 1987 in Payne 1991]; Achagua *kū'wi* ~ *kúwi* MB, *nukuírri* U, *kū'* FZ [Wavrin 1932 in Landaburu 1998-2, Rivet 1948, Meléndez 1990 in Landaburu 1996-2]; Piapoco *cuí* [*kuí*] MB, *nukuíri* my MB, FZH, EF, *cùu* [*kùu*] FZ, *nukuíru* my FZ, MBW, EM [Koch-Grünberg 1923, Wavrin 1932 in Landaburu 1998-2, Wilbert 1966, Klumpp de 1995, Reinoso 1997 in Landaburu 1998-2]; Cabiyarí *aaku* MB, EF [Schauer & Wheeler 1987 in Payne 1991]; Baniwa *nako* ~ *nū'ku* MB, *noko* ~ *nū'ku* A, *nā'ko* EM, *nõkihi* MB, EF, *nõkoiho* ~ FZ, EM [Koch-Grünberg 1911, De Oliveira 1975]; Curripaco *kiiki* ad. MB, EF, FZH, *nukiri* my MB, FZH, EF, *kuuku* or *aakum* ad. FZ, EM, *nukuirru* my FZ, EM [Wilbert 1966, Journet 1993]; Kadaupuritani (Baniwa dial.) *lī'keri* MB, *likeri* EF, *liku̯iru* A, *r(u)ix̯neru* EM [Koch-Grünberg 1911]; Siusi (Baniwa dial.) *nuki* ad. U, *núkeri* ~ *nukiri* MB, EF, *núkuiru* ~ *nukuiro* FZ, EM, *núheri* ~ *nuhéhi* GF [Koch-Grünberg 1911, De Oliveira 1975]; Sucuriu (Baniwa dial.) *nookeri* MB, *nukuirô* FZ [De Oliveira 1975]; Carútana (Baniwa dial.) *nókiri* MB, EF, *nókuiru* A, *nōxneru* EM [Koch-Grünberg 1911]; Jurupari (Baniwa dial.) *nukiri* MB, EF, *nukuiro* FZ, EM [De Oliveira 1975]; Arara (Baniwa dial.) *nõkihi* MB, EF, *nõkoiho* FZ, EM [De Olivera 1971, 1975]; Adzáneni *núkiri* my MB, *nunxấũeri* my EF, *nũ̋nxeru* my EM [Koch-Grünberg 1923, De Oliveira 1971, 1975]; Tariano *nukhí* ad. MB, EF, *nukhíka* my MB, my EF, *núkhiri* ref. MB, EF, *nukhúi* ad. FZ, EM, *nukhúika* my FZ, my EM, *nukwiru* ref. FZ, EM, also *likúiru* A, *likú̯iru* EM, *nõh̯uē'ri* GF [Koch-Grünberg 1911, Aikhenvald 2003]; Guarequena *kā'ka* MB, *nukko* id. [Koch-Grünberg 1911, Créqui-Montfort & Rivet 1922, Rivet 1948]; Bare *nukáka* MB, *nokā'ka* EF, *kō'ko* A, *nakō'tari* A, EM [Koch-Grünberg 1911, Créqui-Montfort & Rivet 1922, Rivet 1948]; Guinaú (dial. of Bare?) *núxku* Oheim (MB), *nú:ku* EF, *naxkú* EM [Koch-Grünberg 1923, Rivet 1948]; Baniva *nokko* ~ *nū'khu* U, *nā'ko* EF [Créqui-Montfort & Rivet 1922, Rivet 1948, Lopes de Sousa 1959]; Yavitero *nō'ko* MB, *iku* U [Koch-Grünberg 1911, Créqui-Montfort & Rivet 1922, 1948, Mosonyi 1987]; Mandawaka *nu-kā'ka*, my MB, my EF, *kaka* U, *nu-kō'ro* my EM [Créqui-Montfort & Rivet 1922, Koch-Grünberg 1923, Rivet 1948]; Manao *ghooko* MB, *nakueru* FZ [Martius 1867b, Goeje 1947, Rivet 1948]; Wainumá *ghoxhoi* MB [Martius 1867b, Rivet 1948]; Waraicú *ghuk* MB, *ghay* GM [Martius 1867a & b, Rivet 1948]; Cawishana *nokodza* FZ [Martius 1867b]; Passé *nokoi* FZ [Martius 1867b]; Mariaté *aku* FZ [Martius 1867b]; Kariaí *nurey* MB [Martius 1867b];

Arawan: ***_koko_ MB, EF** [Dixon 2004b]; Paumarí *koko* MB, EF, FZH *kukŭ* EM [Schuller 1910, Créqui-Montfort & Rivet 1922, Rivet 1948, Odmark & Landin in Merrifield 1985, Dixon 2004b]; Madi *koko* MB, EF [Dixon 2004b]; Jarawara *koko* ad. & ref. MB, EF, FZH [Dixon 2004a & b, Vogel 2006]; Banawá *koko* MB, EF [Dixon 2004b]; Jamamadi *koko* S [Créqui-Montfort & Rivet 1922, Rivet 1948]; Sorowahá *koko* MB, EF [Dixon 2004b]; Dení *koko* ~ *kuku* ad. & ref. MB, EF [Dixon 2004b, Koop 2008]; Kulína *koko* MB, EF, FZH [Adams-Spell & Woods-Townsend 1975, Dixon 2004b]; Madihá (Kurina) *kuku* ad. U, EF, *kukú* compère [Archives Rivet 1923]; Madihá (Zuwiha) *kóko* (♀)HF, *kukú* EF [Archives Rivet 1923]; Colina *kaku* comrade [Archives Rivet 1920];

Aymaran: Jaqaru *kăka* U [Hardman 1981, 1983, Belleza Castro 1995]; Kawki *kaka* U [Hardman 1969];

Barbacoan: Guambiano *kásuko* U [Rivet 1941]; Colorado *ʔa*^{h}*’ko* B [Moore 1966];

Boran: Bora (Miraña) *axa* DH, SW [Guyot 1977];

Cariban: *_koko_ U (MB) [Goeje 1946], **$^{\#}$_kuku_ GM, (FZ, EM?)** [$^{\#}$Matthey & Bancel 2014]; **Parukotoan**: Hixkaryana (Parukoto) *owhoko* U, EF [Derbyshire 1965]; **Pekodian**: Bakairí *kogo* MB, FZH [Oberg 1948b, 1953]; Arara (Pará) *koko* ~ *kɔkɔ* MB, FB, FBS, FZS [Teixera-Pinto 1995, 1997]; **Venezuelan**: Makushi *kooko*, *okoko* GM, *koko* MB [Martius 1867a & b, Arch. Rivet 1921, Williams 1932, Nimuendajú 1955]; Pemón *koʔwai* ad. GM [Thomas 1983]; Taurepang (Pemón dial.) *koʔway* ~ *kokó* (my) GM [Koch-Grünberg 1923, Pessoa 2006]; Arekuna (Pemón dial.) *kɔkɔi* GM [Edwards 1977, Urbina 1983-84]; Akawaio *’aʔai* M [Edwards 1977]; Panare *koko* FF, (♂)FBS, (♀)MZD, (♂)eB, *kokon* yB, young C [Villalón 1978, Mattéi Muller 1994]; Tamanaco *koko*? [von den Steinen 1886]; **Nahukwa**: Kuikúru *kokojó* ref. GM, other fem. relatives, *o’o* ad. GM [Oberg 1953, Dole 1983, Dole 1984, Franchetto 2010]; Kalapalo *okojo*, *oó* GM, *-kuẽgü* GGP [Guerreiro Jr. 2008]; **Guianan**: Carib (Kariña, Galibi) *molekóko* boy, *kah-tobo* MB, WF, (♂)FZH, MBS, FZS [Goeje 1909, Schwerin 1982]; Carib (Cachama) *káx-topo* GF, MB, EF [Schwerin 1983-4]; Carib (Oiapoque) *ka-tobô* MB, MBS, (♀)FZS [Arnaud 1968b]; Carib (Maroni River) *kax-tobo* MB, (♂)MBS, (♀)FZS, *kooki* B, MBS, S, SS [Kloos 1971, Arnaud & Alves 1975]; Ye’kwana (Makiritare) *kóóko* ~ *kóko* ad. GF, FFZS [Wilbert 1958, 1966, Heinen 1983-84]; Wayana (Roucouyenne) *ku-ni* GM, old woman, *kɔ̃nko* FF, MB, EF, (♀)HB, *akon* eB, *kono* WB, HZ [Crevaux *et al.* 1882, Goeje 1909, Goeje 1946, Hurault 1961]; **Taranoan**: ***_kuku_ GM, (FZ, EM)** [Meira 1998]; Pianacoto *ku-ni* GM [Koch-Grünberg 1908]; Akuriyo (Triometesem) *kuku-ni-komo* GM [Goeje 1946]; Trio *kuku* ad. & ref. (♂)GM, (♂)FZ, WM, (♂)MBW, (♀)GM, (♀)FZ, HM, (♀)MBW, *konóka* ref. (♂)FZH, WB, ZH, etc., *koko* ref. & ad. (rarely) (♀)FZD, (♀)MBD, (♀)BW, HZ [Goeje 1909, Rivière 1969, Arnaud & Alves 1975]; Carijona *kuuku* GM, FZ [Wavrin 1932 in Landaburu 1 1996b, Robayo 1989 in Landaburu 1 1996b]; Hianácoto-Umáua *kúúku* GM [Koch-Grünberg 1908]; **Residue:** Apalaí *e-o-ø* my MB [Camargo 2001-2002]; **Unclassified:** Sapara *kuu-nutú* GM [Koch-Grünberg 1923]; Apiaka *koko* U [Ruhlen 1994]; Pimenteira *kuckú* MB, boy [Martius 1867a & b, Goeje 1909, Schuller 1911]; Paravilhana *gocko* MB, *tamuy gocko* GF [Martius 1867b]; *Yukpa:* Japréria *yuvan-koko* A [Oquendo 2004];

Chapacuran: Wari (Pakaasnovos) *xa’* yB, yZ [Vilaça 1995, Everett & Kern 1997]; Itene *uhui* U [Créqui-Montfort & Rivet 1913];

Chibchan: *’_káka_ F, *’_gaka_ EM, *_A’kī‘kī‘_ old, **$^{\#}$_kuku_ MB** [*Constenla Umaña 1981, $^{\#}$Matthey & Bancel 2014]; **Pech:** Paya *ukù* (♂)MB, (♂)EF, *oká* (♀)B, *u’á* (♂)FZ, (♂)EM, *kaiká* (♂)ZD [Schuller 1928, Conzemius 1928]; **Core Chibchan: Votic:** Rama *kā‘lŭŋ* (♀)MB [Lehmann 1914, Ortiz & Rigby 1992]; **Isthmic:** Cabécar *kaga* F, *kĕ́gölö* [kígɷlɷ] MB, *kë‘kë‘wa* [kíkíwa] anciano(a), *ǰak* EM [Lehmann 1920, Schuller 1928, Stone 1962, Margery Peña 2004]; Cabécar-Čiripo *kaka* F, *keke* señor [Lehmann 1920, Schuller 1928]; Cabécar-Estrella *káka* F, *keke* señor [Lehmann 1920]; Cabécar (Tukurrike) *ka* F [Lehmann 1920, Rivet 1949]; Bribri *akë‘këla* [akígıla] anciano, *ǰak* EM, *wökela* [wɷkela] GF [Margery Peña 1982]; Boruca *sugu* anciano, viejo, *kakat* B [Pittier 1941, Rivet 1949]; Teribe *kok* F, *kéga* U, *kégi* EF, *k*^{h}*òkí* GF, *k*^{h}*égɛ́* MB, *k*^{h}*ɛ́gɛ̀* EF [Lehmann 1920, Rivet 1949, Oakes 2001]; Teraba *kok* F, *kega* U, *kegi* ~ *keke* EF, *keh* señor [Lehmann 1920]; Dorasque (Changuena) *kága* F [Rivet 1949, Pinart 1890]; Dorasque (Chumulu) *wóka* GF [Pinart 1890]; Dorasque (Gualaca) *óka* GF [Pinart 1890]; Cuna-Tule *ansakka* EF, *nukkwa* anciana [Prince 1913, Wavrin 1932 in Landaburu 4 1999b]; San Blas Cuna *sakka* EF, *mukkwa* GPZ [Stout 1947]; **Magdalenic:** Chibcha (Muisca) *kaka* GM, *guaka* (♀)EF, (♀)DH [González de Pérez 1987]; Tunebo (U’wa) *kaka* GM, FFZ, MMB [Osborn 1995, Headland 1997]; Tunebo (Sínsiga of Cobugon) *kaka* GM [Rivet 1924]; Cágaba (Cogui) *na-kukui* A, *na-kagi* ~ *no’kh̲agi* ~ *gag*i EM, *na-s̲akha* GF, *na-saka* ~ *saxa* GM [Lehmann 1920, Wavrin

1932 in Landaburu 2 1998, Gawthorne 1985, Ortiz 1996 in Landaburu 2 1998]; Damana (Guamaca Malayo) *anz-kugu ~ kúgu* U, GF, *anzi-gwagi* A, *anzg̱aga ~ sáhga* GM [Lehmann 1920, Schuller 1928, Wavrin 1932 in Landaburu 2 1998, Trillos 1996 in Landaburu 2 1998]; Atanques (Cankuamo) *kúku* U, *káke* F, *gwasi* A, *sukui* GM [Celedon 1892 in Landaburu 2 1998, Lehman 1920, Rivet 1924, Schuller 1928, Wavrin 1932 in Landaburu 2 1998]; Arhuaco (Bíntukua, Ica) *te̱gwe* U, *kákə ~ káke* F, *-gwati* A, *nə-takə* GM [Lehmann 1920, Wavrin 1932 in Landaburu 2 1998]; **Unclassified:** Nutabé *guaku* old woman [Rivet 1949];

Chipaya-Uru: Chipaya *ouiké* GCh [Snethlage & Snethlage 1932];

Chocoan: Emberá Catio *káu* D [Wavrin 1934 in Landaburu 4 1999b]; Choco (Emberá) *kau* D [Faron 1961, Reichel-Dolmatoff 1962]; Epena *'kʰau* D, girl [Quiro Dura *et al.* 2007]; Noanamá *ka* D, *kúi* Z, *kawá* EM [Reichel-Dolmatoff 1962];

Cholonan: Hibito *kotk* F, *keek* M [Rivet 1949];

Chonan: Tehuelche *koka* F, dueño, *kok'an* madrina, *ko:n* U, *goo-* B, Z, *qon* GM [Fernandez Garay 2004]; Selknam *ho'o* GF, *hoho'nh* GM [Lowie 1933]; Puelche *ukɨčɨ* B [Viegas Barros & Casamiquela 2007];

Guahiboan: ***axu* MB, EF, *akwe GM** [Matthey & Bancel 2014]; Guahibo (Hiwi, Sikuani) *axuyo ~aáxuyo ~ ahuyo* MB, EF, FZH, *akúyo* EF, *akue̱* U, *ákwe* GM, *noko* my U [Melgarejo 1886, Rivet 1948, Wilbert 1966, Metzger & Morey 1983, Ortiz Gomez 1983, Merrifield & Kondo 1985, Queixalos 1989]; Cuiva (Hiwi) *ahuyo ~ axujo* MB, FZH, EF, *akwe* GM [Arcand 1976, Ortiz Gomez 1983]; Guayabero *kuewon* B, *kuewow* Z [Tobar Ortíz 1989]; Macagua *aknú* U, *akúi* ad. EF, *akuéy* GM [Buenaventura 1993, Dixon 2004b];

Jêan: Panará (Kren Akarore) *kokripia* EM [Giraldin 1994, 1997]; **Southern Jê **kãke* eB, **kakrã* MB, EF** [Jolkesky 2010]; Kaingang (Toldo das Lontras, Nonoai, etc.) *kakre* MF, MB, WF [Baldus 1952, Hicks 1971]; Kaingang (Rio Grande do Sul) *kakrë* MF, MB, EF, *kɛ̃ke* eB, eZ [Magno de Aquino 2008]; Kaingang (Palmas, Paraná) *kakrá ~ kakrẽ̱* EF, *kãke* eB, eZ [Loukotka 1929, Baldus 1935, Wiesemann 2001]; Kaingang (Serra do Chagú, Paraná) *ikãke* B [Hanke 1947]; Kaingang (Duque de Caxias) = Xokleng *kɔkla* ceremonial F = MB, *kaké* relative, cousin, rarely A [Henry 1964]; Ingáin *kau* U [Ambrosetti 1896];

Jivaroan: Jivaro Achuar *kai* (♀)B, (♀)Z [Taylor 1998]; Huambisa *kair* (♀)Z [Jakway 1987]; Aguaruna *kaig* (♀)Z [Bant 1994];

Kamakanan: Kamakan *gköong* GF, MB [Martius 1867b, Ignace 1912, Silva Martins 2007];

Karajá language area: Karajá *hi* (♀)eB, *ixi* yB [Krause 1911, Pétesch 2000];

Karirian: **$^{\#}$*kuku* MB** [Matthey & Bancel 2014]; Kipea-*kukú* U [Mamiani 1698, 1699, Adam 1897, Goeje 1932, Rodrigues 1948]; Dzubukua *-kuku* U [Bernardo de Nantes 1709, Correia de Queiroz 2008]; Pedra Branca *cuccuh* MB [Martius 1867a & b, Adam 1897, Rodrigues 1948]; Sabuya *cuccùh ~ cuccu* MB [Martius 1867a & b, Adam 1897, Rodrigues 1948]; *Krenákan:* Botocudo *ñ-gikę˚n* F, (♀)EF [Nimuendajú 1946];

Lenca-Misumalpan: Lenca: **koko* GF, (EF) [Constenla Umaña 2002]; Lenca (Honduras) *koko ~ kogo* EF [Lehmann 1920, Schuller 1928]; Lenca (Chilanga - El Salvador) *koko ~ kogo* GF, anciano, *koh* U, eB [Lehmann 1920, Schuller 1928, Campbell 1976, Del Rio Urrutia 2004]; **Misumalpan: **kukuŋ -ki* GF, (Un)** [Constenla Umaña 2002]; Miskito *kuki-ki* my GM, *kū'ka* GM, anciana [Heath 1913, Lehmann 1920, Espinoza 2001]; Cacaopera *kuku-* GF, GM [Noguera 1855 in Lehmann 1920]; Matagalpa *kuku-ke* my U [Campbell 1975]; Ulwa (of Karawala) *ku:kuŋ̊-ki* my GF [Green 1999]; Ulwa (Pearl Lagoon: Kukra?) *kukuŋ* GF [Lehmann 1920]; Taüaxka-Ulwa *kobunh-ki ~ kokohu(ni)* GF [Lehmann 1920]; Sumo *kukō'-kē* my cousin, *ku:kuŋ-ke* my GF [Lehmann 1920]; Sumo-Panamahka *ku:kuŋ* GF [Green 1999]; Sumo (von Houwald) *kukuŋ-* GF [von Houwald 1980]; Tuahka *koko* GF [Lehmann 1920];

Lule-Vilelan: Lule-Tonocote *kue* FZ [Machoni de Cerdeña 1878];

Makúan: Puinave *hika* A, *-u'ɨi* EF [Koch-Grünberg 1923, Wavrin 1932 in Landaburu 2 1998, Giron 1993 in Landaburu 2 1998, Richardson 2007]; Hupda *ʔuw* GF, *ʔɔh* GM, *ʔoʔ* eB, *ʔũʔ* Sib [Erickson & Erickson 1993]; Nadëb *óów* GF, *hooñ* GM [Weir 1984];

Mascoyan: Enlhet dial. of Sanapaná *haawok eŋak* eB [Unruh & Kalisch 1997]; Mascoy (Western) *koko* MB, *hawok* eB [Braunstein 1983]; Mascoy (Eastern) *hawok* eB [Braunstein 1983];

Matacoan: Chulupí (Nivaclé) *k'ucaʔx* old man, *k'ucxaʔ* old woman, *yi-kakt'eč* my EF, *yi-kakt'e* my EM [Wicke & Chase-Sardi 1969, Seelwische 1980, 1990, Braunstein 1983, Renshaw 2002]; Chorote *kihil*, *kíhi* GF, old man [Karsten 1932, Braunstein 1983]; Maca *xuhkew* eB, *k'ucaχ* old man, *-kewket* EF, *-kewket-iʔ* EM [Braunstein 1983, Gerzenstein 1994, 2007]; Mataco (Wichí Lhamtès Vejoz) *-kó* M, *kawaklá* ZH [Hunt 1940, Braunstein 1983];

Maxakalian: Maxakali *xuxyã* MB, EF, FF, MF, GGF, *xukux*, FZ, EM, GM, GGM [Popovich 1980];

Muran: Pirahã *kaai* D [Everett 2005];

Nambikuáran: ***#kũ:nkV-* MB, EF, FZH, GF** [Matthey & Bancel 2014]; Halotesu *-kṵ̃ka-nu-su* MB, EF [Kroeker 1996]; Kokozê = Juina Kitaulhu *kôkô-zu* U [Roquette-Pinto 2005 (1917)]; Waklitisu *-kun̥ku-nú-su* MB [Boglar 1960]; Elotasu *-kun̥ku-nú-su* MB [Boglar 1960]; Mamaindê *k^{h}ṳh-ni-ɾu* MB [Lévi-Strauss 1948, Eberhard 2009]; Tawandê, Latundê *-kún-de* MF, EF [Lévi-Strauss 1948]; Sabanê *koóka ~ kooko* MB, EF, WB, FF, FM [Lévi-Strauss 1948, Antunes de Araujo 2004];

Paezan: Paez *ká^{h}ka* MB, MZH [Wavrin 1931 in Landaburu 3 1999a, Bernal Villa 1955, Rojas Curieux 1995 in Landaburu 3 1999a];

Pano-Tacanan: **kuku* ~ **kuka* MB, (EF) [*Girard 1971]; **Panoan**: ****koka* MB, EF**, ****koko* nephew** [Shell 1965, Girard 1971]; Group I: Amahuaca *koká* MB , (♀)F [Kensinger 1995]; Group II: Cashibo *kuku* MB, EF, FZH [Kensinger 1995, Wistrand-Robinson 1998]; Shipibo-Cunibo (Tschama) *koka ~ kú:ka* MB, WF, *koko*, (♀)ZS, ?(♀)BS [Tessmann 1929, Shell 1965, Loriot *et al.* 1993, Kensinger 1995]; Capanahua *koka* U, EF [Kensinger 1995, Loos 2003]; Panobo (Huariapano, Pano) *koka* U [Parker 1992]; Pano (Navarro) *kuka* MB, HF, *kuku* sobrina [Navarro (d'Ans 1970)]; Group III: Isconahua *koka* MB, EF [Niemeyer Cesarino 2008]; Cashinahua *kuka* MB, FZH, EF [Kensinger 1995]; Catuquina *koka* MB [Coffaci de Lima 1997, Kennell Jr. 2007]; Marubo *koka* (♂)eMB, (♂)eZS, (♀)MB, (♀)ZS [Kensinger 1995]; Sharanahua *koka* MB [Kensinger 1995]; Yaminahua *koka* ad. MB, EF, FZH, ad. (♂)ZS, ad. (♀)S [Townsley 1994]; Shanenahua *kuka* U [Vieira Cândido 2004]; Yawanahua *kuka* U [Santos de Paula 2004]; Marunahua *kóká* U [Key 1968]; Kaxariri *ku'ku* U, EF [Calvacante Souza 2004]; Poyanahua *kuka* ad. U [Santos de Paula 1992]; Group IV: Korubo *koko* MB [Oliveira 2009]; Mayoruna (Tabatinga) cucu [*kuku*] U [Martius 1867b]; Matses *kukú* MyB, EF, yZS, *koka* MB, *kako* (♀)BS [Fields & Merrifield 1980, Erikson 1986, 1994, Fleck 2003, 2005]; Matis *kuku* MyB, EF, FZH, yZS [Kensinger 1995, Erikson 1999, Vicente Ferreira 2005]; Chacobo *koko* U [Key 1968]; Pacahuara *koko ~ kuko* U (MB), EF, sobrino [Orbigny 1839, Rivet 1910, Créqui-Montfort & Rivet 1922]; **Tacanan**: ****kuku* U** [Girard 1971]; Ese Ejja *toto*, *tsotso* MB, *kaka* F [Pitkin Wyma 1962, Shoemaker 1963-75, Key 1968, Chavama 2003, Lepri 2005]; Huarayo *toto* MB, *kaka* FB [Chavarria Mendoza 1984]; Araona *huhu* MB [Brinton 1892, Créqui-Montfort & Rivet 1922, Key 1968, Pitman de 1981]; Tacana *huhu*, *xuxu*, *hoho* MB, U [Brinton 1892, Créqui-Montfort & Rivet 1922, Schuller 1922, Key 1968]; Reyesano *hʋhʋ* U, *koʔa* M [Key 1968, Caceres Raldes n.d.]; Cavineña *kuku*, *ekuku* (my) MB [Brinton 1892, Créqui-Montfort & Rivet 1922, Key 1968, Camp & Liccardi 1989, Guillaume 2008];

Qawasqaran: Alakaluf (Qawasqar) *xoyko-las* old woman [Clairis & Viegas Barros 2007];

Quechuan: **kaka* MB, WF, WB, ((♂)MBS) [Matthey & Bancel 2014]; Cuzco (1586, 1607 etc.: Q2) *caca* [kaka] U (honorific) ; *caca(y)* [kakaj] MB, WF, WB [Barzana 1586, Holguín 1607, Yaya 2008]; Coastal Peru near Lima (1560: Q1?, Q2?) *caca(y)* [kakaj] (my) MB, WF, WGF [Santo Tomas 1560]; Central Sierra east of Lima (1618: Q1) *caca* [kaka] WF, U [Arriaga 1621]; Modern Q'ero (Q2) *kakay* my MB, MBS [Webster 1977]; Pitumarca (Q2) *kaka* MB, HB, WB [Milicic 2011]; Apurimac Quechua (Q2) *kaka* MB, WF [Camacho *et al.* 2007]; Ayacucho Quechua (Q2) *kaka* MB [Zariquiey & Córdova 2008, Soto Ruiz n.d.];

Sáliban: Sáliba *koko* man, *tihoho* A, *kaku* Z [Martius 1867b, Tastevin 1922 & Estrada 1993 in Landaburu 2 1998];

Timotean: Timote-Kuika *kuxioy* GM [Rivet 1926]; Mucuchí *kaak* man [Rivet 1926];

Tukanoan: Coreguaje *cu'eʉ* GF [Gralow & Merrifield 1985]; Siona *k^{w}ɨ̃* MB, MBS, MBSS [Vickers 1989]; Secoya *k^{w}ɨ̃* MB, MBS, MBSS [Vickers 1989]; Orejón (Mai Huna) *gũ* MB, *yeke* gũ FZ [Bellier in Santos & Barclay 1994]; Barasana *gagü* eB, *gago* eZ [Hugh-Jones 1979, Chemela 1996];

Tupian: Western Tupian: **Arikem**: Arikém *u-kërą* (♂)B, *u-isįką* my B (ego ♀), my Z (ego ♀) [Nimuendajú 1932]; Karitiána *syky* (♀)B [Landin 2005]; **Mondéan**: ***#ko(t)kó(t)* ad. MB, (EF, GF)** [Matthey & Bancel 2014]; Suruí *kokó* ad. U (MB) [Bontkes & Merrifield 1985]; Sanamaicá (Mondé, Salamãy) *koko* U (MB) [Greenberg & Ruhlen 2007]; Cinta Larga *kokó* ad. MB, EF, WB, ZH, MBS, *kyptï* (♂)B [Dal Poz Neto 2004]; Zoró *kutkut* MB, FZ, GP [Dal Poz Neto 2009]; Gavião *kótkóòt* ad. MB [Moore pers. com.]; **Purubora:** Puruborá *kɔkɔ* MB [Galucio 2005]; **Ramaráman**: Káro (Uruku) *iká* eZ, *hakóiga* H [Schultz 1955, Gabas Jr. 1989]; **Tuparían**: Tuparí *-ike* (♂)eB [Caspar 1975]; **Eastern Tupian**: **Awetí**: Awetí *i-kywýt* (♀)B [Galvão 1953, Zarur 1975]; **Mawé** Mawé (Maue) *uy-kiwid* (♀)B, *ohẹkẹ́ẹ* B [Nimuendajú 1929, Koch-Grünberg 1932]; **Mundurukún:** Mundurukú *ukipid* my eB (ego ♂), *o-köt-köt* yMZS, yFBS [Nimuendajú 1932, Murphy 1956]; **Tupí-Guaranían: *#kiwyra* (♀)B** [Matthey & Bancel 2014]; Guayakí (Aché) *kyvangi* (♀)B [Clastres 1968]; Chiriguano (Guarani) *kiuỹ̆*, *kigwi* ~ *cekigwi* (♀)(e)B [Koch-Grünberg 1902, Braunstein 1983]; Shetá *koti* B [Loukotka 1929]; Sirionó, *áke* SS, DD [Scheffler & Lounsbury 1971, Califano 1999]; Cocama-Cocamilla *kį̇wį̇rą* (♀)B [Espinosa 1935]; Tupinambá *kybyra* (♂)B [Fernandes 1963]; Tenetehára (Guajajara) *he-kiwyra* (♀)B [Wagley & Galvão 1949]; Parakanã *-kywyra* (♀)B [Fausto 1995]; Tapirapé *che-kyvuyra* (♀)B [Baldus 1970]; Araweté *čiwi* (♀)B [Viveiros de Castro 1992]; Kayabí *kɨwɨt* (♀)B [Weiss 1985]; Apiaká *erarkuireè* B [Koch-Grünberg 1902]; Kagwahív (Tenharim) *kuvyr* (♀)B [Betts 1981, Kracke 1984]; Kamayurá *ie-kywýt* (♀)B [Oberg 1953, Galvão 1953]; Urubú-Kaapor *ihẽ̆kywyr* (♀)B [Kakumasu 2007]; Oyampi *kakay* eB [Hurault 1962];

Witotoan: Nonuya *hōkhá* man, *hohēkŏ* WB, HB, *hokohokŏ* WZ, HZ [Rivet & Wavrin 1953]; Ocaina *hoh́ə̆* man, *hăhāhō(h)ə̆* WB, HB, *kōhōhŏ* cousin [Rivet & Wavrin 1953]; Witoto *mota uaikika* GF (F + old man) [Murdock 1936];

Zamucoan: Ayoreo *axai* (♀)B, *axu* MZ [Braunstein 1983]; Chamacoco *-okok* EF [Baldus 1932]; Ebidoso *p-o̲ho̲t'* my EF [Baldus 1932]; Tumereha *p-óhó'* my EF [Baldus 1932];

❖ Isolates

Awaké (Sape) *ma-kohaí* ~ *ma-ku̲haí* my GF, *kohai* old man [Koch-Grünberg 1923, 1928];

Camsa *-ki* MB [Matteson 1972];

Canichana *eu-axa* my GF [Créqui-Montfort & Rivet 1913];

Cayuvava *-kiče* U [Key 1975];

Cofan *tõʔto* U [Borman 1976];

Kapixaná (Kanoe) *kű̃kű̃ì* man, *keke* GM, *ūkū* WBD, HBD [Nimuendajú 1928, Archives Rivet, Bacelar 2004];

Kwaza *ha'kai* GP [Van der Voort n.d. (2004)];

Mapudungu (Mapuche, Araucano) *hueku* ~ *weku* MB, *kacü* FZH, *kuku* FM, FFZ [Latcham 1904, Faron 1956];

Masaca (Aikana) *kokomai* U [Ruhlen 1994];

Mochica (Chimu) *kokəʊd* (♂)eZ, (♂)A, *ikiş* EM [Salas 2002];

Movima *aiku* A, *ákai* eB, eZ [Créqui-Montfort & Rivet 1914];

Munichi *tcʔaʔa* GF [Gibson 1996 (2008)];

Oti (Chavante?) *koaka* B [Ruhlen 1994];

Taushiro *'ukku* (♂)Z [Ortiz 1975];

Trumai *kokó* MZ, FBW, *aoké* EM, FZ [Galvão 1953, Sutherland Louis 1971];

Urarina (Simacu) *ka-kaun* FZ (dubious) [Walker 2009];

Warao *ku*, *da-ku* MB [Williams 1928, Heimen 1997];

Yaté (Fulnio) *ī-xi* B [Pinto 1956].

MOTHER TONGUE
Journal of the Association for the Study of Language in Prehistory • **Issue XIX • 2014**

Indo-European and Dravidian: Some Considerations[1]

Stephan Hillyer Levitt

Abstract
This paper examines several considerations with regard to Indo-European and Dravidian from the vantage of Nostratic in the light of one another, and in the light of in certain instances comparable features in other language families grouped under the rubric Nostratic. The paper suggests that Nostratic languages spread with the original spread of anatomically modern humans from Africa, placing the linguistic data in sync with recent genetic and climatological studies. The linguistic features considered are the Indo-European and Kartvelian mobile *s*-, certain sound correspondences in Dravidian (*c: t*, *r: l, ṉ, ṟ* and *l: ḷ, l: t, ḷ: ṭ*) and their Nostratic parallels and implications, metathesis in suggested cognates between Indo-European and Dravidian and its Nostratic implications, inserted and dropped *r* in Indo-European etyma from the vantage of Nostratic. In addition, a few individual Indo-European and Dravidian etyma are considered that indicate the antiquity of the Dravidian forms within Nostratic.

1. Introduction

It is becoming clear that the similarities between lexical items in different language families are less a matter of chance than of a relationship as sure as that of different lexical items within the different languages of a single language family, such as Indo-European (IE).

Most recently, such relationship has given rise to such lexicons of suggested correspondences as Bomhard (2008 and 2011) and Dolgopolsky (2008). Neither Bomhard nor Dolgopolsky have to date, though, made use of my articles relating to aspects of the Nostratic hypothesis, nor the work of South Indian linguists, in general, with regard to Nostratics.

Just as when Sir William Jones turned to Persian and Sanskrit, he was able to see the connection between different branches of IE; so, when we turn to Tamil, the most conservative of the Dravidian languages, with a classical literature going back to the early centuries BCE, we are able to see the connection between different families of languages – pointing to a monogenesis of language. It is to this that I attribute the independent observations of G. Devaneyan and myself that Dravidian is related to IE; and within IE, that it is related most closely to Germanic.

---

[1] I would like to thank Bob Scott, currently Co-Director of Columbia University Library's Digital Humanities Center for his help in accessing information in Bomhard (2008 [ebook]). I would also like to thank the anonymous reader for *Mother Tongue* for his worthwhile suggestion. An earlier version of this paper, with major printing errors in the Bibliography, appeared in *International Journal of Dravidian Linguistics* 42.2 (June 2013): 63-99. The customized version of the TransIndic Transliterator font used for some of the diacritics in this paper is available from www.linguistsoftware.com/tintu.htm, +1-425-775-1130.

Devaneyan (1902-1981) was a linguistically savvy Tamil scholar whose Nostratic thesis is often distorted and misunderstood in the West. His arguments go back, though, to such pioneers in Dravidian studies as R. Caldwell (1814-1891) and G. U. Pope (1820-1908), and for instance to D. Savariroyan (fl. 1899-1914) and S. Gnana Prakasar (1875-1947).

In 1998 and 2000 I published two papers in *The Journal of Indo-European Studies* on the possible relationship between IE and Dravidian. With regard to my filtering of Savariroyan's, K.C.A. Gnana Giri Nadar's and Devaneyan's suggested Nostratic correspondences between Tamil and IE in the 2000 article, Edgar Polomé commented that "I may not agree with some of your etymologies but I find them quite challenging. ... Congratulations on a nice piece of work!" (undated correspondence from July 1999). I note, I now would accept more of Devaneyan's suggested correspondences than I did at that time.

I followed through with additional considerations and correspondences in *Folia Linguistica Historica* in Levitt (2003), in *International Journal of Dravidian Linguistics* in Levitt (2007a, 2010, and 2011-12), and in two short articles in *Chemmozhi* in Levitt (2007b and 2008). See also Levitt (2012) in *Mother Tongue*.

My method, by and large, is to keep semantic transparency paramount, and to see if there are logical connections between forms that can be argued on this basis, using attested sound correspondences. There appears to be a stability of meaning in a large number of items.

It is my contention that many of the sound changes and alternations that appear in Dravidian are very ancient and can be seen in Nostratic in general, no doubt from pre-Dravidian.

It has been argued by some that there have been untold diffusions of peoples and tongues, mergings, and fresh dispersals. Because of this, we cannot be certain at deep time depths of the actual relation – genetic, borrowing, or simple chance resemblance after attrition and/or innovation. Thus, Hungarian *ház* 'house' looks very like Germanic versions of 'house', but it is seen to be cognate with Finnish *kota* 'tent, cloth hut', and to go back to Finno-Ugric **kota* 'tent, hut, house'. So also, *devil*, *diable*, *diablo* etc. in other languages suggest perhaps genetic cognates but documentation says they are all loans and/or devolutes from Gk. *diabolos*, literally 'slanderer'. Many suggested cognations at deep time depths may be right, but how do we verify this? In the absence of documentation, how can we be sure?

And, in fact, such indeed is so. Yet, by keeping meaning constant and staying within the parameters of clear sound alternations we can see such developments.

For instance, *DEDR* 1796 Ta. *kurai* carries such meanings as 'to bark, jubilate, shout', 'noise, roar, shout' and 'dog'. Also compare *DEDR* 1901 Ta. *kūraṉ*, which forms mean 'dog' and 'bitch', and *DEDR* 2122 Ma. *kora* 'asthma'. These forms give rise to Skt. *kurkura*, *kukkura*, *kukura* 'dog', and a variant *kukkuṭa* (*DEDR* 1796). Regarding the Sanskrit and related Indo-Aryan words *CDIAL* 3329 etc. comments "Onom[atopoetic]". Levitt (1998: 151-152 [no. 23]) has argued these are loan forms in Sanskrit (see also *DEDR* 1796), related on a deep level to Lat. *canis*, Eng. *hound* and, Levitt has argued, through metathesis, Eng. *dog*. The genetically related Sanskrit cognate is *śvan*.

With regard to Eng. *dog*, see the Hindi form *kuttā* 'dog', *kuttī* 'bitch', listed in *CDIAL* 3275 and cross-referenced there with Skt. *kurkura*.

To be kept in mind here is that there is an alternation in Dravidian between *r* and *l* (see Section 2.3). And an alternation between *l* and *n* in Dravidian forms as against IE forms can be seen in a number of correspondences argued in Levitt (1998). Levitt (2003) greatly simplifies the etymology of Eng. *dog* by demonstrating that in metathetical forms in Dravidian medial or final *l* or *ḷ* becomes initial *t*- (see Levitt 2003: 179).

In the present context, probably related in English is the English word *growl*, which is imitative according to *Webster's* 508b, and the related Scottish word *gurl* (see *OED* 6: 896a, 963c). Such may be a reduplicated onomatopoetic form, or a form showing an echo (as also OE. *docga*). And as well no doubt related are Eng. *cur*, ME. *curdogge*, ME. *curren* 'to growl' and Germ. obs. and dial. *kurren* 'to growl, grumble, murmur, coo' (see *Webster's* 278b and *OED* 4: 135c for these and related forms, and their history).

The explanation of how we can have forms of such different shapes as *hound*, *dog*, *cur*, and *growl* that stand side-by-side in English all coming from the same protoform **ku-r/l/n/t* is that they are from different dialectical sources. The generic name for 'dog' in Old English and the Teutonic languages generally, was *hund* (*hound*), pre-Teut. **kun-*. *Dog* appears in Old English with its origin and previous history unknown. *Cur* appears first in Middle English and is of continental origin. *Growl*, with but two exceptions, appears first in the 17th c. Its continuity from the early usages is questionable, though it may have been preserved in some dialect. Its 13th c. usage is in Anglo-French, and has an ultimately Teutonic source.

Further explaining such variation from a single original source standing side-by-side in English, except for *hound*, which as a substantive is now used in specialized usage or fashion, the other words seem to appear originally in specialized usages (*OED* 4: 135c, 921b, 6: 896a, 963c, 7: 432bc).[2]

While I think that such work as Bomhard's and Dolgopolsky's, V. Illič-Svityč's earlier work, and J. Greenberg's work on Eurasiatic, which set up clear-cut sound correspondences is premature, and that we ought to be casting a broader net as we are still in the stage of collecting data, I do think that they are very helpful starts. Thus, Illič-Svityč's examples of correspondences between Nostratic *l*- and Dravidian *t*-, which Bomhard has not adopted, however, and which Dolgopolsky seems to have separated from one another (no. 1281) or just discarded as not proved by other etymologies (no. 1309), for example, was expanded by me in Levitt (2003). On this basis I was able to suggest still other correspondences between IE and Dravidian, and support further such formative processes I had given example of in Levitt (1998) as metathesis.[3]

---

[2] In this connection, I might point to a recent article by Sayers (2008) in which the English forms *dog* and *cur* are explained with what I would judge to be very forced arguments. The proposal here is much simpler and more straightforward, and is therefore preferable.

[3] Compare Illič-Svityč's correspondences as reported in Levitt (2003: 175) with, for instance, Bomhard's listings under Nostratic **t'uk'-*, **t'ok'-*, and Nostratic **lak*^{h}-, **lək*^{h}-, accessed by an "All Fields" search for 'beat' and 'leg' respectively in the "Proto-Nostratic Dictionary" in Bomhard (2008[ebook]); and Dolgopolsky's discussion of the Dravidian data at the end of his entry no. 1281 for Nostratic 'push, shove, thrust, fasten to' and his comment at the end of entry no. 1309 for Nostratic 'to bend' in Dolgopolsky (2008). On account of the difficulty in reproducing Dolgopolsky's reconstructed forms here, I omit them.

Further, neither Bomhard nor Dolgopolsky take into account euphonic combination, which I hope Levitt (2010) has shown is indeed in force in Nostratic.

I propose here to treat several topics, in the course of which I will bring in some of the material I have treated earlier, as well as new material.

First off, though, something ought to be said about the spread of our Nostratic languages and their dating.

There are various "popular" scenarios regarding the origin of Proto-IE (PIE). Two predominate in recent years. One, advanced by Colin Renfrew, argues for the spread of IE languages with the spread of agriculture from southern Anatolia starting in the 8th millennium BCE. The other, advanced by Marija Gimbutas, argues that it was the spread of the Kurgan people from their homeland in the river basins of the Volga and Don in eastern Europe that led to the dispersal of IE languages from the early 5th millennium BCE onward. The International Genomic Project under the direction of Luigi Luca Cavalli-Sforza has combined the two, arguing for primary and secondary diffusion (Haarmann 1998: 391-393).

Still others, such as Paul Thieme and Ward H. Goodenough, have argued for a north Central European homeland, specifically the North European Plain of Germany and Poland, "or at least the eastern end in Poland and the western Ukraine" (Goodenough 1970: 254, 262).

Many Russian archeologists now consider the Kurgan culture to be Indo-Iranian. And, indeed, the contact of this culture with a culture seen by modern Russian archeologists to be the precursor of Baltic and Slavonic culture, would explain the observed linguistic contact between Indo-Iranian and Balto-Slavonic (Kuznetsov 2005: 325; Telegin 2005: 339; Burrow 1973: 18-23). This argument is more in keeping with the spread of Nostratic languages as I see it.

I hold a position which, while it sounds strange when one first hears it – as it did to me, makes perfect sense on reflection, and fits with the genetic, climatological, and geographic considerations.

I would see Nostratic as going back to the original peopling of the world by anatomically modern man (AMH). This opinion is bolstered by linguists of Australian aboriginal languages, who see these as having hived off from Dravidian, or pre-Dravidian, if you would (Dixon 1980: 236-237). So, also, Trombetti (1906-20). Recently, Masica (2001: 255), too, has opined on the basis of parallels with Australian aboriginal languages that a pre-Dravidian substratum in South Asia may well go back to the original peopling of the world by AMH.

Linguistically, the evidence includes phonological features such as both retroflexes and alveolars beside dentals, and typological similarity such as word order, an agglutinative morphology, and an inclusive/exclusive distinction in non-singular first person pronouns, to note just three such similarities.

There is also a connection between Australian aborigines and Dravidians in the kinship system, and in the use of the boomerang.

Ramanathan (1983-84, 2009a; "Introduction" to Devaneyan 2004: 17^1-20^1) has proposed a number of cognates between Australian aboriginal languages, just as Blažek (1988, 2006) has proposed cognates between Australian aboriginal languages and Dravidian. Blažek, though, views these as loan forms in Dravidian from an earlier population, as he accepts Zvelebil's 1972 model of the descent of Dravidian into South

Asia through the northwest, and later dating for Nostratic than I would argue for (see Morris 2006: 99).

Note that the Y-chromosome genetic marker M130 is found with frequency among Australian aboriginal populations and Dravidian populations in South India, but is rare elsewhere. And recently, an entire village of Kallar near Madurai in Tamil Nadu, with which caste group the boomerang is particularly associated, was found to have the Y-chromosome genetic marker M130 (Wood 2007: 15-16; National Geographic Society and IBM 2005-2006: video and pamphlet, p. 7; Oppenheimer 2007[2004]: 171, 185, 187, and esp. chart on 186).

AMH appears to have reached South Asia by way of a coastal route before the eruption of the Toba volcano in Sumatra in 74,000 BCE. A “furrow” in the genetic evidence in South Asia argues for this (Oppenheimer 2007[2004]: 82). Also, recent archeological research has found similar sets of stone tools in South Indian sites both below and above the layer of ash from the Toba volcano, giving archeological foundation to the genetic data (Patel 2007: 15).

From there, AMH went to Australia perhaps by 65,000-70,000 BCE or earlier, according to Oppenheimer (2007[2004]: 159-163, 192), at least by 40,000 or 50,000 BCE according to others (Dixon 2002: 8-9), being cut off from the rest of mankind by the rising oceans since about 6,000 BCE.

Kivisild et al. (2003: 313) argue that there has been little genetic influx into South Asia from external regions since the original peopling of South Asia by AMH. Further, as Oppenheimer (2007[2004]: 152) has argued with regard to the Y-chromosome mutation M17 (R1a), often referred to as the “Caucasoid” genetic marker, it appears to have developed in South Asia, where it is most diverse (see also Sahoo et al. 2006: 843a).[4]

Both on climatological and geographical grounds, when AMH left Africa, he could not have gone through the Middle East, but would have gone by the coastal route (Field and Lahr 2005, Oppenheimer 2007[2004]: 68, 78-83).

AMH would have moved out of southern Asia to the region of the Zagros mountains, western Iran and eastern Turkey, and from there he would have moved further into the Eurasian landmass, and into the Levant and North Africa, after 40,000-50,000 BCE with the retreat of the glaciers, etc. (Oppenheimer 2007[2004]: 129-131, 133, 138-141, see map on 132), a second migration into Eurasia from southern Asia coming around 33,500 BCE by one of two possible routes – a trans-Caucasus route, or by way of Kashmir through the Central Asian steppe (Oppenheimer 2007[2004]: 144-146, 149-150, see map on 137).

This data fits with the linguistic scenario of Afroasiatic breaking off from Nostratic first, then Dravidian, then Kartvelian (Bomhard 2008[ebook]: Ch. 1, 2011: 4-5).

2.1. *Mobile s-*

In IE, there is an initial *s-* that appears inexplicably in front of some forms only without any regularity in a set of etyma, forming in those instances an initial consonant group. Some examples are Skt. *tanyati* ‘thunders’, Lat. *tonāre*: Skt. *stanayitnu* ‘thunder’, see also Gk. *steno*, OChSl. *stenjǫ* etc.; Skt. *tṛ-*, *tārā* ‘star’: Skt. *stṛ-*, Av. *star*, Gk. *aster* etc.;

---

[4] For fuller discussion of this data, see Levitt (2009: 140-144, 2010: 53-55, 57-61).

Skt. *phāla* 'ploughshare': Pers. *supār*; Skt. *lavaṇa* 'salty, salt': see Lat. *sal*; Skt. *nihākā* 'fog', *nīhāra* 'mist, dew': *snih-* 'to be moist' etc.

Burrow (1955: 80, 1973: 81) is of the opinion that the phenomenon is due to loss of initial *s-*, and is most probably the result of some kind of external euphonic combination affecting initial *s-* in the IE period.

Vladislav Illič-Svityč was of the opinion that it developed only in Nostratic words with an inherited internal palatal element (Dolgopolsky 2008: 18).

The interpretation of Dolgopolsky (2008: 18, 19) is that Nostratic emphatic consonants in word initial position were pronounced in IE with an initial pre-aspiration. Usually, this pre-aspiration left no trace. But in roots with an inherited palatal element the whole root was palatalized, and the pre-aspiration was transformed into a kind of **h* that later yielded a movable **s-*.

Dolgopolsky also mentions a typological parallel in the Turkic language Salar. In Salar, the vowel *i* before voiceless *p*, *t*, and *k* (fortes) was transformed into *i* + a pre-consonantal sibilant.

Neither Bomhard (2008, 2011) nor Dolgopolsky (2008) reconstruct a mobile *s-* for Nostratic. See Bomhard (2011: 78 [no. 143] and 206 [no. 595]) and Dolgopolsky (2008: nos. 309, 362, 364, 417, 439).

There is a mobile *s-* as well in Kartvelian languages. Here, it is in both consonant clusters and before vowels. Thus, Georg. *ni(s)ḳarṭ-* 'beak': Svan *niḳraṭ*, *niḳräṭ* (Common Kartvelian **ni-(s)ḳrṭ-*); Georg. *sadgis-* 'awl': Megr. *odgiš-* (Georgian-Zan **(s)a-dg-is*$_1$*-*); Georg. *savarʒel-* 'arm-chair': Laz *orʒu-*, *orʒo-*, *orzo-* 'bench, chair' (Georgian-Zan **(s)a-warʒ-el-* 'seat'); Georg. *saxl-* 'house': Megr. *oxor-* 'dwelling, abode; farmstead', Laz *oxor-* 'house' (Georgian-Zan **(s)a-xl-* 'house') (Klimov 1998: 125, 143, 172-178, 338; see esp. notes under Common Kartvelian **pu-* 'to chop, hack, cut to pieces' [p. 152] and Common Kartvelian **(s)ʒie-* 'milk' [pp. 177-178]).

Also, Ohno (1983) reports doublets with and without initial *s-* in Japanese.

In this context, we ought perhaps to reconsider Proto-Dravidian **c-*.

Proto-Dravidian **c-* can be analyzed as having undergone several different treatments in the various Dravidian languages, which treatments may represent different phonemes in the protolanguage; or rather, it can be analyzed as having undergone several irregular and incomplete sound changes (Krishnamurti 2003: 121; Emeneau 1988: 247a-260b, 265a-266b).

This phoneme is represented in some Dravidian languages by *c-* and in others by *s-*. In some languages, there is free variation. Toda represents it always by *t-*. Some dialects of Gondi, Kuwi, and sometimes Brahui change it to *h-* (Krishnamurti 1961: 8).

By one of these sound changes, it is by present analysis lost in South Dravidian, Telugu, and some dialects of Gondi (Krishnamurti 2003: 121-124; Emeneau 1988: 248a-250b; Burrow 1947). Thus,

(1) Ta. *il* 'non-existence, death', *illai* 'it is not (in classical Tamil with a complete neg. paradigm); Ma. *illa* 'does not exist, there is not; no, not'; Ka. *illa*, *illaṃ*, *illavu* 'is or are not; no, not'; Te. *lē-* '(complete neg. paradigm, 3 sg. non-m. *lēdu*) 'be not'; Pa. *cila* 'does not exist'; Go. *sile*, *sille*, *hille*, *hile*, *ille* 'not, no'; Koṇḍa *sil-*, Pe. *hil-*, (in songs) *sil-*, Kui *siḍa* 'not to be'; Kuwi *hill* 'to be not' (*DEDR* 2559) [PDr **cil-*].

(2) a. Ta. Ma. Ka. Tu. Te. *uppu*, Kol. *sup*, Nk. *supp*, Pa. *cup*, Ga. *sup*, *cuppu* 'salt'; b. Ta. *uvar* 'to taste saltish, be brackish'; Ma. *uvar* 'salt taste', *ōr* 'saltness'; Ka. *ogar* 'an

astringent taste'; Tu. _ubaru̥_, _ubāru̥_, _ogaru̥_ 'brackishness'; Te. _ogaru_, _vagaru_ 'astringent taste'; Go. _sawwor_, _sauwar_, _havar_, _hovar_, _ovar_, _ovor_; Koṇḍa _soru_ 'salt' (_DEDR_ 2674[a] and [b]) [PDr *_cup_: *_cuv-ar_ (< *_cup-ar_)].

Ramaswami Aiyar (1929a: 152 [sec. vi], 1929b: 7n., 1930: 171, 1932: 29) views some such alternations of initial *_c_- (Kui _s_-) with ϕ as being a development from an original prothetic front on-glide appearing before original palatal initial vowels. This opinion is consonant with Illič-Svityč's analysis of the IE mobile _s_-. Burrow (1947: 135) dismisses Ramaswami Aiyar's suggestion, just as he dismisses explanation for the IE mobile _s_- that would regard the _s_- as the remains of some kind of prefix.

Devaneyan (1966: 96-99, 2004: 26^1-29^1) takes issue with Burrow's consideration only of aphetic omission, and not the prothetic addition of initial consonants. He notes that various words in Dravidian add prothetic _c-/s-_, such as Ta. _uruḷ_ 'to roll, tumble over and over, become round, revolve, whirl' (_DEDR_ 664[a]): Ta. _curi_, _curuḷ_ 'to be spiral as a conch, whirl round, curl, lie in a circle' (_DEDR_ 2684) and Ta. _ur̤al_ 'to oscillate, swing, be in motion, whirl, revolve': Ta. _cur̤aṅku_, _cur̤al_ 'to whirl, spin, rotate, roll' (both listed under initial *_c_ in _DEDR_ 2698[b] and [a] respectively). In a 1966 article in Tamil in _Centamil̲c Celvi_ he notes _uppu_ = 'effervescent saline soil, or a lump or pinch of salt', _umaṇ_ = 'saline soil, fuller's earth, salt' (_maṇ_ = 'soil, ground, earth' [Fabricius 1933: 770a]), _uvar_ = 'salt which rises up and forms a film at the top of saline fields, saline land'. He comments that the Tamil demonstrative _u_ is the ultimate base of thousands of words indicating forward and upward movement, among others, and _uppu_ and _uvar_ are among those words (Devaneyan 2004: 29^1).

We can also point out that _TED_ 3/2: 55b derives Ta. _cimir̤_ 'to bind', for instance, from Ta. _imir̤_, both listed under *_c_ in _DEDR_ 2542; and _TED_ 1/2: 292b gives as cognates for Ta. _ippi_ 'pearl-oyster, shell': Ka. _cippi_, Tu. _cippi_, Te. _cippa_ (and in Indo-Aryan, Pkt. _sippi_), all also listed under *c in _DEDR_ 2535.

Given the existence of a mobile _s_- in Kartvelian as well as IE, Kartvelian having broken off from Nostratic immediately after Dravidian by present opinion (Bomhard 2008[ebook]: Ch. 1, 2011: 4-5), it makes perfect sense that Dravidian, too, would have a mobile _s_-, and that such forms as in _DEDR_ 2559 and _DEDR_ 2674 ought to be reconstructed as *_(c)il_- and *_(c)up_: *_(c)uv-ar_ [< *_(c)up-ar_]. Note in this regard that in Dravidian, there are no initial consonant clusters in the protolanguage.

That the etyma in _DEDR_ 2559 ought to be reconstructed as *_(c)il_- is also indicated by a suggested correspondence in Levitt (1998), which correspondence was argued by Devaneyan before as well, between these forms and Eng. _no_, _not_ (also _non_-, _un_-, _in_-, _il_- [before _l_], _im_- [before _m_ or _p_], _ir_- [before _r_]), Germ. _nein_, _nicht_, Skt. _na_, _an_- (Pokorny 2005[1959-69]/1: 756-758, PIE *_nĕ_, *_nē_, *_nai_ [*_n̥_-, word negation], see also 1: 17, PIE *_aiu̯_, *_ai̯u_ and Pei 1962: 154-155; Levitt 1998: 148 [no. 17], 2000: 421). The appearance of _n_ in IE for _l_ in Dravidian occurs in a number of other proposed cognates as, for instance, Eng. _inn_: Ta. _il_ 'house, home' (_DEDR_ 494; Levitt 1998: 147-148 [no. 16], see also entry nos. 13, 18, 23 and the comment toward the end of entry no. 13 with regard to Prakrit sound changes). A form with initial *_s_- (PDr *_c_- as presently configured) appears most probably in Lat. _sine_ 'without', Sp. _sin_, Fr. _sans_, Skt. _sanitur_ 'besides, without (with prec. acc.)' (Pokorny 2005[1959-69]/1: 907, PIE *_seni_-, *_senu_, [*s_e_ni_-], *_sn̥-ter_-, see also 1: 318, PIE *_ĕneu_, *_ĕnu_; Ernout and Meillet 1985: 627-628; Levitt 2000: 429 [no. 3]).

Another example in which a prothetic *s-* has been added in IE is Lat. *sōlus* 'alone, sole', Eng. *solo*. This word at present is without convincing etymology (see Pokorny 2005[1959-69]/1: 884; Ernout and Meillet 1985: 634b, Levitt 2012: 112, 160). Levitt would see it as being related to Dravidian words for 'one, single' Ta. *oru*, *ōr* … Ta. *oṉṟu*, *oṇṇu* (*DEDR* 990[a] – [d]). Both Levitt and earlier David (1966: 240, 284-285 [Supplementary Note 4]) see this as related within Dravidian to a root *ol* 'to unite (often with reality), to be joined, possible, feasible' (*DEDR* 1006 Ta. *ollu*, the latter with an enunciative vowel), which underlies such forms in Tamil as *okka* 'together' and *oṉṟu* 'one' (*DEDR* 990[b] and [d] respectively; see Levitt 2012: 104-106). It is these forms as a group that Levitt sees as being related to the IE words for 'one' such as Eng. *one*, Germ. *ein*, Lat. *ūnus* (Pokorny 2005[1959-69]/1: 286, PIE **oi-nos*; Levitt 1998: 144-147 [no. 13], 2012: 108-109, 149, 151; see also Devaneyan 1966: 232).

Still a third form can be found in a proposed relationship between Ta. *el* 'sun', also 'lustre, splendor, light, daytime' (*DEDR* 829): Eng. *sun*, Lat. *sol*, Gk. *hēlios* (Devaneyan 1966: 232; Levitt 2000: 424, 2007b: 20b, see also 2010: 72-74).

Also compare Ta. *nākam* 'snake' < Ta. *nakar* 'to creep, crawl': Eng. *snake*, OE. *snaca* < OE. *snican* 'to creep, crawl' (Devaneyan 1966: 250, Fabricius 1933: 594b-595a, 578a). In Finnish we also have *naakia* 'to hunt crawling on the ground'.[5] And in Hebrew we have *nāḥaš* 'serpent', noted in Cohen (2010: 898a) under *ḤNŚ*, which latter carries among its meanings 'to hunt', 'to seduce', and 'reptile'. Both Dravidian forms are omitted from *DEDR* as well as *DBIA*. Burrow and Emeneau, though, no doubt consider Ta. *nākam* 'snake' to be a loanword in Dravidian from Skt. *nāga* 'snake'. Sanskrit, though, does not show a corresponding verbal form, as we have in Tamil. And within Dravidian, we have other related forms which are reflected in Ta. *nā*, *nākku*, *nāvu* 'tongue', *nāvu* 'to thrust out the tongue and move it sideways, mock by thrusting out the tongue' (*DEDR* 3633); Ta. *nakku* 'to lick, lap' (*DEDR* 3570); Ta. *naku* 'to laugh, smile', *nakkal* 'laughing, mockery' (*DEDR* 3569). The relationship between the etyma in *DEDR* 3633 and Ta. *nākam* 'snake' would be that a snake constantly thrusts its tongue in and out of its mouth sideways.[6]

In Levitt (2007b: 20b-21a), it was argued because of these points that Skt. *nāga* was rather a loan form from Dravidian in Sanskrit, the parallel forms in IE displaying a prothetic *s-*. In light of the IE mobile *s-* phenomenon, though, the Sanskrit form may well be inherited within IE. We just cannot be sure in such situations. In that Sanskrit does not display the matching verbal form or related forms containing any of the related cognitions that can be found in Dravidian (except, perhaps, Skt. *nagara* 'town, city' < *DEDR* 3568 Ta. *nakar* 'house, mansion, temple, palace, town, city', judged by Burrow and Emeneau to be a loan form in Sanskrit) does however suggest that perhaps Skt. *nāga* is indeed, nevertheless, a loan in Sanskrit.

These examples (and there are more) of prothetic *s-* in proposed IE cognates with Dravidian are, no doubt, related to the phenomenon of a mobile *s-*.

With regard to a mobile *s-* being found in both Kartvelian and by the argument here, in Dravidian, and with regard to Kartvelian having separated from Nostratic immediately after Dravidian by present theorization, note that Fähnrich (1965) discusses extensive similarities between Kartvelian and Dravidian of both phonetic and grammatical nature, to which Fähnrich added extensive lexical similarities. To be kept in mind here is that Fähnrich (1965) assumes an "Ibero-Caucasian" protolanguage that includes Kartvelian and North Caucasian languages. While Nostraticists today include Kartvelian in Nostratic, together with Dravidian and Afroasiatic, the North Caucasian languages are not so included.

---

[5] I would like to thank Prof. Panu Hakola for bringing my attention to the Finnish form (correspondence dated 17 December, 2011).

[6] Compare Eng. *snicker*, considered by *Webster's* 1099b to be "imit[ative]". So also *OED* 15: 850b, which cross-references Eng. *nicker* (*OED* 10: 394b) and Eng. *snigger* (*OED* 15: 853c), both of which it also considers to be "imitative". Especially note that in Eng. *nicker* we have a form without the initial *s-*.

As to phonetic similarities, Fähnrich notes (pp. 138-139) that the phonetic systems of the individual Kartvelian and Dravidian languages moved away from one another considerably over time. Nevertheless, certain general characteristics can be laid down. For instance, in comparison with the Kartvelian languages, the Dravidian languages show relatively more vowels and relatively fewer consonants. Further, the Kartvelian languages developed glottalized consonants, which the Dravidian languages lack. The Dravidian languages, on the other hand, possess retroflex consonants, which are foreign to the Kartvelian languages. But against these differences, there exists nevertheless a whole lot of common ground. For instance, all consonants of both language families have points of articulation similarly organized according to the following groups: guttural plosive, labial plosive, dental plosive, dental fricative, *m*, *n*, *r*(*l*).

The syllable structure, Fähnrich (1965: 139-140) notes, as well shows considerable common ground between the two language families. By far the most frequent initial sound is the guttural plosive. Further, as an initial, *r*(*l*) is most frequently replaced.

With regard to word construction, in both language families the word, root, and syllable were originally one and the same. The word was monosyllabic, being the pure root, and as yet no formatives were added. Further, in both language families, the basic models for word construction are similar, the most important structures being CVC (doubled, CVCCVC), CVCV, CVCVC [C = consonant, V = vowel].

As to grammatical similarities, Fähnrich points out (p. 141) that with regard to typology, the primary principle of word derivation is in both language families agglutinative. Further, both language families prefer to represent grammatical references with synthetic types of expression.

As to rules of structure (pp. 141-143), he provides an extensive listing of similarities. These include his noting that the usual sequence of parts of the sentence is subject, object, predicate. The synthetic character of the two families brings with it that this sequence can be changed when emphasis is demanded on a part of the sentence, or on account of some other circumstance. The verb is marked differently in transitive and intransitive in both language families. In both language families, the conjugation system is greatly developed. The infinitive functions in a number of languages as a noun (Georgian, Malayalam, Telugu, Kui, Brahui).

Fähnrich also provides an extensive listing of morphological similarities (pp. 144-146). For instance, a dental element is used to indicate the past tense (Tamil, Malayalam, Parji, Kui, Georgian). The infinitive ends with '*a*' (Kui, Georgian, Hunzib, Bats). A verbal suffix, formed with a vowel + '*n*', appears in Kota, Tulu, Georgian, Laz, Lezgi, Avar, Lak, Chechen, Adyghe, and Kabardian. Another verbal suffix, '*p*/*b*/*v*', is found in Georgian and many Dravidian languages. A genitive suffix '*-n*' is found in Tamil, Kannada, Telugu, Parji, Gondi, Brahui, Chechen, Tabasaran, and Lezgi.

In Fähnrich's listing of similarities between the two language families in both rules of structure and in morphology, the sheer number of parallels noted presents a very strong argument for genetic relationship.

With regard to lexical similarities (pp. 146-158) interestingly he lists correspondences between Kartvelian and Dravidian in words for 'one', 'two', 'six', 'seven', 'eight', and 'not to be', 'to be', 'fire', and a wide assortment of basic vocabulary. This list, it is noted, was to be continued. It does not cover all possible initials, such as guttural plosives or labial plosives. I am not aware of such a continuation having been published. The lexical correspondences offered, though, provide further compelling argument for genetic connection between the two language families.

Fähnrich (1965: 136-137) places the relationship between Kartvelian and Dravidian in the context of a hypothesized Asian language family most of which languages, such as Sumerian, Elamite, Hurrian, Hattic (the so-called Proto-Hittite), Urartian, etc., are dead. In the case of Dravidian, though, the closest language of which is Brahui, we still have today living languages. As to the geographical distance between Kartvelian and most Dravidian, he points to the distance between Kartvelian and Basque, which he sees to be related to Kartvelian, the two being separated by Indo-European languages as also Dravidian and Kartvelian.

With regard to Hurrian [= Mitanni] and Dravidian, see Brown (1930). With regard to Elamite and Dravidian, see McAlpin (1981). See also Ramanathan (2009b) for a review of Szalek (2006) regarding Lycian and Lydian, and regarding the Pelasgian language. According to Herodotus, the Lycians (Herodotus' 'Termilai') came to Anatolia from Crete.

Consonant with this, in a 1981 paper on Kartvelian and Sumerian, Fähnrich argues (pp. 90b-91a) that the Sumerians were not autochthonous in southern Mesopotamia, and that a Dravidian substrate can be found in Sumerian. In this regard, Fähnrich lists 17 lexical items in Sumerian which do not allow Sumerian etymologies, but for which Dravidian etymologies can be found. These include (1) Sum. *Uri* 'city in Sumer': Ta. *ūr* 'village, city', Ma. *ūr* 'village, city' … Te. *ūru* 'village, city', Naiki *ūr* 'village' and (17) Sum. *ud* 'storm': Ta. *ūtai* 'wind', Ma. *ūttu* 'to blow, to blow about' … Te. *ūdu* 'to blow about', Gondi *udānā* 'to blow about'. Fähnrich argues that before the Sumerians arrived in southern Mesopotamia, Dravidian speech was current there.

In a more recent article, Fähnrich (1991) provides 90 regular phonemic correspondences between the protolanguages reconstructed for Kartvelian and Dravidian based on 129 lexical correspondences. In his listing of phonemic correspondences he notes the number of instances of each correspondence in the body of lexical evidence he has pulled together; and in his listing of lexical correspondences he refers back to the number he has assigned to each phonemic correspondence suggested for each phoneme of each lexical item.

He argues that the number of parallels between the Dravidian and Kartvelian protolanguages makes the assumption of an accidental similarity very improbable. He adds that the nature of the parallels – the semantic range of the protolanguages' lexicons, the elementary area of the vocabulary covered, and above all the root morphemes concerned, also excludes the borrowing of one protolanguage by the other or borrowing from a third source by both protolanguages. Therefore, a genetic connection between the two protolanguages is suggested, if not proven.

It thus ought not be surprising that we would find examples of mobile *s*- in both Kartvelian and Dravidian, as well as in Indo-European, Japanese, and the Turkic Salar, and a suggestion of same in Hebrew *nāḥaš* 'serpent' (i.e., 'snake') and its associated root *ḤNŚ*.

2.2. *c: t*

One of the treatments of **c*- in Dravidian is its sporadic development to *t*-. This is attributed to the postulated original nature of **c*- as an affricate (Emeneau 1988: 250b-151a; Krishnamurti 1961: 12; Burrow 1947: 142-143). And there is, indeed, indication of an affricate pronunciation of *c*- in the South Asian area, outside of the evidence from Dravidian, in for instance the Sanskrit verbal form *ścut-/cyut-/*often in the later language, *ścyut*- 'to flow, trickle, ooze' (*Ṛgveda* +; *KEWA* 3: 380, *EWA* 2: 658-659 note, "without certain connection" [translation mine]). Emeneau (1988: 252b-255b) lists sets of etyma in Dravidian in which (a) **c*- is sporadically replaced by *t*-, and in which (b) **c*- is sporadically replaced by *t*-, with occurrences of *ϕ* in South Dravidian and Telugu.

As noted earlier, it is my contention that many of the sound processes we observe in Dravidian are also applicable to Nostratic (see, for instance, Levitt 2010). In Levitt (2000: 430 [no. 5]), an example was given of this with regard to this phonological development. This was a proposed genetic relationship between Eng. *tell*, *talk*, OHG. *zellen* (Pokorny 2005[1959-69]/1: 193, PIE **del*-; Buck 1949: 18.21; Pei 1962: 45-46) with Ta. *col* 'to say, speak, tell, mention, utter, express, recite, repeat; n. word, term, saying, speech, proverb, praise', *collavu* 'saying, telling, proverb'; Ma. *colluka*, *celluka* 'to say, declare, order'; Ka. *sol*, *sollu* 'to say, speak, tell; n. word'; Pa. *cul*- 'to say'; etc. (*DEDR* 2855). With regard to the Germanic forms, Buck (1949: 18.21[5]) notes "outside root connections … dub[ious]".

Emeneau (1988: 265a-266b) also lists examples in *DEDR* of the sporadic replacement of **t*- by **c*- (or as Burrow 1947 referred to it, original *c-/s*-). These are offered by Emeneau without phonological explanation (p. 265a). To these, we may also add a connection to a Sanskrit form appearing first in Vedic literature, Skt. *śūrpa* 'a winnowing basket or fan (i.e., a kind of wicker receptacle which, when shaken about, serves as a fan for winnowing corn)' and Ta. *tūr̲r̲u* 'to scatter, winnow, throw up as dust in the air'; Ka. *tūr̲u* 'to winnow, drive off chaff from grain by means of the wind'; Tu. *tūpuni* 'to winnow, fan'; Te. *tūru* 'husks of grain', *tūru-paṭṭu* 'to winnow', *tūrpiḍi* 'winnowing'; Kol. *tūrpet*- 'to winnow' (*DEDR* 3402). This connection, first proposed by Burrow (1945: 118 [no. 48]), is noted in *DEDR* with question; but in this context it appears to me, to be fairly certain. The loan would be from Central Dravidian. Such can be determined by the shape of some of the Central Dravidian forms in comparison with those in South Dravidian.

In the context of the common occurrence of *s*- for **c*- in Dravidian, and my contention that the Dravidian phonological developments are ancient and can be seen as going back to pre-Dravidian, we can also suggest that it is this replacement of **t*- by *c-/s*- that is behind the IE demonstrative pronoun **to*-, **tā*-, **ti̯o*- (Pokorny 2005[1959-69]/1: 1086-1087) appearing in the Nom. sg. m. and f. as **so(s)*, **sā* (Pokorny 2005[1959-69]/1: 978-979). Such forms are also listed separately by Dolgopolsky (2008: nos. 2310, 2006) and Bomhard (2011: 68 [no. 114], 95 [no. 201]).

2.3. *r: l, ṟ, ṉ and l: ḷ, l: t, ḷ: ṭ*

It is often said today that the alternation of *r* and *l* is not a feature of Dravidian, though such an alternation was observed early by, for instance, Konow (1903: 456), who cites in that place Robert Caldwell, as well.

l develops to *r*, of course, in Vedic Sanskrit as against Iranian. And in eastern Prakrits such as Māgadhī, *r* becomes *l*, always in Māgadhī.

It is my contention that *r* and *l* alternate in forms in Dravidian, but is most often masked by allophones of *l* such as *ṟ* and *ṉ*, which appear for instance in euphonic combination in classical Tamil. Thus *kal* + *pu* > *kaṟpu*, *kaṭal* + *karai* > *kaṭaṟkarai*, *kaṭal* + *paṭappai* > *kaṭaṟpaṭappai*, *cel* + *mati* > *ceṉmati*, *kal* + *mār* > *kaṉmār*, *vēl* + *miḷai* > *vēṉmiḷai*, *vel* + *vēl* > *veṉvēl* (Rajam 1992: 105). There is evidence that such rules of euphonic development follow phonological developments within Dravidian (Krishnamurti 2003: 93-94; Zvelebil 1970: 172, 178-180; Ramaswami Aiyar 1935-38/n.s. 28[1]: 20-21; see Levitt 2010: 64-70). It is this which leads Emeneau (1967[1957]: 141-142) to see evidence for an alternation between *r* and *ṟ*, and "morphological doublets" with the alveolar nasal *ṉ* and alveolar *ṟ*.

An alternation between *r* and *l* as such can be seen in colloquial Tamil, which contains such alternation probably dialectically as well in some colloquial forms for literary Tamil forms, and in loanwords. See, for instance, Ta. *pantar* 'a thatched shed': *pantal* (so, also, Ka. *pandar*, *pandara*: *pandal*, Te. *pandiri*: *pandili*; *DEDR* 3922) and Ta. *cāmpar* 'ashes': *cāmpal* (*DEDR* 2453). For literary Tamil *lāntar*: *lāntal* (< Eng. *lantern*), colloquial Tamil in Tirunelveli district has *rāndal*. In literary Tamil such alternation can be seen in, for instance, Ta. *mūtari* 'to establish with evidence, confront with proof': *mūtali* (see also Te. *mudarakincu* 'to remind of something painful or humiliating, ask, question': *mudalakincu*; *DEDR* 5040) and in Ta. *tōl* 'to be defeated, lose; n. defeat', *tōlvi*: *tōr*, *tōrvai* (*DEDR* 3558). There is, as well, the standard development in Tulu both medially and finally of **l* > *r*. With regard to these points, see Zvelebil (1970: 144, 143) and, for a fuller discussion, Levitt (1989).[7] Also, see *TED* 12: 24ab [2.141] which gives 7 examples of *r* > *l* (4) and *l* > *r* (3) in Tamil etymology.

There does not appear to be a geographical pattern in Dravidian with regard to the alternation of *r* and *l*, as we have in Indo-Aryan.

With regard to *ṟ* and *ṉ*, Devaneyan (1966: 204) is of the opinion that these are late developments, and that before the formation of alveolar *ṉ*, the dental *n* was medial and final, as well as initial. It is only Tamil and Malayalam that show evidence for alveolar *ṉ* beside dental *n*. Zvelebil (1970: 130) charts their distribution in Tamil. Zvelebil (1970: 129-130), too, is of the opinion that they represent one single phoneme **n* in Proto-Dravidian.

An alternation between *l* and *n* in proposed IE cognates with Dravidian has been indicated in Section 2.1.

Hodge (1998) has noted that there also appears to be alternation without obvious pattern in Afroasiatic between *r*, *l*, and *n*, though he tries to develop one from his data.

---

[7] Levitt (1989) contains many typesetting errors. It is hoped that these will be corrected if I am able to publish a collection of my papers on Dravidian topics in the near future.

Pokorny (1960: 161) also pointed to an alternation in Afroasiatic between *r*, *l*, *n*, and *d* as well.

With regard to an alternation between *l* and *t* in Dravidian, Levitt (2003) picked up on an earlier observation of Vladislav Illič-Svityč that Nostratic *l-* corresponds to Dravidian *t-* in initial position (see Illič-Svityč 1989: 151 and 1990: entry nos. 255, 260, 261). Levitt's jumping off point was correspondences which this observation led to regarding Nostratic ***lakʌ* 'leg' and Dravidian **tāk* 'walk' (*DEDR* 3151) with a correspondence he had made earlier (1998: 139 [no. 5]) with PDr **kāl* 'leg, foot' (*DEDR* 1479); and regarding PDr **āḷ* 'person' and related forms (*DEDR* 399, 400), and the etyma in *DEDR* 3196 Ta. *tāṉ* 'oneself' and *DEDR* 3162 Ta. *tām* 'they, themselves'. This observation that medial and final **-l*/**-ḷ* alternate within Dravidian with forms with initial **t-* leads to a large number of interesting correspondences, such as Ta. *el* 'lustre, splendor, light, sun, daytime' (*DEDR* 829): Ta. *tī*, *tīy* … n. 'fire, lamp, heat'; Ma. *tī* 'fire'; Te. *tīṇḍra* 'light, brightness, heat', *tīṇḍramu* 'heat, light, splendor' (*DEDR* 3266) and Ta. *teṟal* 'heat'; Nk. *tirup* 'sun's ray'; Go. *ter-* 'to be fierce (heat of the sun)', *taritānā* 'to be hot (of sun)', *teṟk-* 'to warm oneself by fire'; Koṇḍa *teṟvel* 'sunshine' (*DEDR* 3440); compare Eng. *sun* (see Section 2.1) and Skt. *div*, *dyu*, *deva* (see Section 2.6). Also, Ta. *kāl* 'air, wind' (*DEDR* 1481), Ka. *gāḷi*, *gāḷ* 'wind, air' (*DEDR* 1499): Kur. *tākā* 'air, wind, breeze' (*DEDR* 3149); compare Eng. *gale* etc. And, Ka. *mēlamba* 'the black humble bee' (*DEDR* 5098): Ta. *tīm* 'sweet', *tīyam* 'sweetness', *tēm* 'sweetness, honey, honey-bee' (*DEDR* 3268[a]); Ta. *tēṉ* 'honey', *tēṉ-ī* 'honey-bee', *tīn-toṭai* 'beehive'; Tu. *tiga* 'beehive, honey'; Te. *tēne* 'honey', *tē̃ṭi* 'a large black humble bee' (*DEDR* 3268[b]); compare PIE **médhu* (see below, and Section 2.6).

It does not seem possible to tell which is primary, forms with **-l*/**-ḷ* or **t-*.

The correspondences gathered in Levitt (2003) appear to suggest a split in Dravidian, or pre-Dravidian, between **l* and **ḷ*. So also, Ta. *kalakala* 'to rustle, tinkle, rattle': *kaḷakaḷa* 'to rattle, chatter, gurgle' (*DEDR* 1302); and Ta. *kol* 'to kill, murder'; Br. *xalling* 'to strike, kill' (*DEDR* 2132): Ta. *koḷ* 'to strike, hurt', *kōḷ* 'killing, murder' (*DEDR* 2152).[*]

---

[*] P. Ramanathan has communicated to me (correspondence dated July 15, 2013) that the gloss 'killing, murder' for Ta. *kōḷ* in *DEDR* comes ultimately from a gloss for the word by Swami Nathaiyar in his edition of the *Jīvacintāmaṇi* (2nd ed., 1907; 3rd ed., 1922). This was picked up by Madras University's *Tamil Lexicon* (6 vols., 1928-39) and was from that, transferred by Burrow and Emeneau to their *Dravidian Etymological Dictionary* (1961) and to *DEDR*. P. Ramanathan notes, though, that in its context, in stanza 264 of the *Namakaḷ Ilampakam*, the traitor Kattiyaṅkaraṉ enters Jīvakaṉ's palace with the objective [*kōḷ*] of murdering Jīvakaṉ, Ta. *kōḷ* meaning 'objective' and only metaphorically, in this particular context, 'killing'. P. Ramanathan does not think that the form conveys the meaning 'killing' anywhere else in Tamil literature. It would thus fall in with *DEDR* 2151 Ta. *koḷ* (*koḷv-*, *koṇṭ-*) … *koḷai* … 'determination'; *kōḷ* 'taking, receiving, accepting, holding …'; the latter also carrying the meanings 'decision, determination, conclusion'.

Fabricius (1933: 308b-309a [*koḷ*], 320b [*kōḷ* = verbal noun]) lists both the transitive meanings and the intransitive meanings, listed separately in *DEDR* 2151 and 2152 respectively, in the same entries.

Note in this regard Nk. *goṛ-* (*goṭ-*) 'to strike, beat, KILL' in *DEDR* 2152. Also, in *DEDR* 2152, Ma. *kōḷ* 'hitting, wound, damage'; Te. (B.) *konu* 'to be pierced as by an arrow'; Kol. *go·l-* (*goḍḍ-*) 'to beat, shoot with bow', (P.) *goḍ-* 'to cut with axe'. It was speculated in Levitt (2003: 188, under entry no. 47) that both *DEDR* 2152 and *DEDR* 2132 might be related on a Nostratic level to Uralic **lükkʌ-* 'pierce, insert, push' and Altaic **lükä-* 'pierce', as given by Illič-Svityč (1990: entry no. 261). Also note here Eng. *kill*, OE. *cwellan*, caus. of OE. *cwelan* 'die', ME. *culle*, *kille*, earlier also 'strike, beat, knock' (see Levitt 1998: 139-140, entry no. 7). [SHL]

There is as well an alternation of medial and final **l*/*ḷ* and medial and final **t* in Dravidian, but it does not appear to be as thoroughgoing.

Thus, for instance, a connection appears to maintain between Ta. *kutirai* etc. 'horse' (*DEDR* 1711[a]) and Ta. *kul*^{1} 'to move forward, to go forward, to move the body in a graceful and affected manner' (as does a horse) (*TED* 2.2: 506a). The standard traditional explanantion of Ta. *kutirai* is that it is derived from Ta. *kuti* 'to jump' (Devaneyan 1966: 58; Zvelebil 1970: 100; *TED* 2.2: 414a). Note, though, the etymologies given in *TED* 2.2: 412a for *kuti*^{1} 'jump' < *kutu*, but for *kuti*^{4} n. 'endeavor' < *kutu* < *kul*. Compare Eng. *colt* 'the young of a horse', of obscure origin according to *OED* 3: 505b.

So also, *DEDR* 2231 Ta. *kōru*, *kōlu* 'to bale, draw up (as with an ola basket)' at present lists with question Koḍ. *ko·t-* 'to fill by scooping'.

One reason for the sparsity of such data would appear to be the aforementioned split in Dravidian, or pre-Dravidian, between **l* and **ḷ*, and the alternation across the board in Dravidian between *ḷ* and *ṭ*. Thus,

(1) Ta. *kuḷam* 'tank, reservoir, lake'; Te. *kolanu*, *kolāku* 'tank, pond'; Go. *kol* 'tank'; Kui *glūnju* 'a small pool, puddle' (*DEDR* 1828): Ta. Ma. *kuṭam* 'water-pot'; Ta. *kuṭukkai* 'coconut or other hard shell used as vessel, pitcher'; Ma. *kuṭukku* 'shells (as gourds) used as vessels'; Te. *kuḍaka*, *kuḍuka* 'cup, bowl, any cup-like thing' (*DEDR* 1651).

(2) Ta. *il* 'house, home, PLACE' (*DEDR* 494): Ta. *iṭai* 'middle in space or time'; Ka. *eḍa*, *eḍe* 'place or time between, interval, distance, middle'; Te. *eḍa* 'space, interval, distance', *eḍamu* 'intervening space, distance' (*DEDR* 448). See *TED* 1.2: 265ab, 260b-261a, which comments *il* → *iḷ* → *iṭu* → *iṭai*. Note, further, in *DEDR* 494 Ta. *il*, Kui *iḍu* 'house, dwelling, shed, hut'.

(3) So also, Ta. *cuṭu* 'to be hot, burn'; Te. *cū̃ḍu* 'to burn, brand with a hot iron or the like'; Konḍa *suṟ-* 'to roast, burn (incense)'; Kur. *kuṛnā* 'to grow warm, become hot, be heated, cook on live embers, bake on an open fire' (*DEDR* 2654). *TED* 3.2: 212ab derives Ta. *cuṭu* from *cuḷ*, which carries a meaning 'pungency, acrimony', and which in turn *TED* 3.2: 271b derives from *cul* (see also the etyma on 272ab for the semantics here).

TED 12: 24b [2.143, 144] gives 3 examples of *l* > *ḷ* within Tamil etymology, and one example of *ḷ* > *l*. *TED* 12: 19a [2.84] gives two examples of *ḷ* > *ṭ*.*

In this regard, David (1966: 274) has pointed out that when a word occurs in two forms in Tamil, with both -*ḷ* and -*ṭ*- in its basic root, the form with -*ḷ* is always the earlier form.

---

* The forms cited in *TED* for *l* > *ḷ* are: Ta. *aṭṭalai* 'apartment on an upper storey; watchtower on a fort; raised covered platform from which one keeps watch on a garden, a field, a village, etc." [*TED* 1/1: 102b] – Ta. *aṭṭaḷai* 'upper storey; covered platform for watching a garden' [*TED* 1/1: 103a]. Ta. *alapputal* 'to chatter, prattle, gossip, talk in vain, to talk nonsense, to talk incoherently' [*TED* 1/1: 434b] – Ta. *aḷapputal* 'chatter' [*TED* 1/1: 502b]. Ta. *alamaral* 'whirling, spinning around; confusion, perturbation, sorrow, fear, dread' [*TED* 1/1: 436a] – Ta. *aḷamaral* [*aḷamarutal* 'to be bewildered, to be distressed' (*TED* 1/1: 504b)]. The forms cited for *ḷ* > *l* are: Ta. *poḷḷāmaṇi* (= *poḷ* + *ā* + *maṇi*), *TED* 6/3: 216b 'see *pollāmaṇi*' [*poḷḷal* 'boring a hole; chiseling, as a stone; hole, rent, fissure, puncture; …' (*TED* 6/3: 216ab)] – Ta. *pollāmaṇi* 'gem which has not been bored; flawless gem; god, as immaculate' [*TED* 6/3: 206a]. My focus it must be emphasized is Dravidian, or pre-Dravidian, and not Tamil as such, in which P. Ramanathan has rightly pointed out to me *l* and *ḷ* are not allophones of one another (correspondences dated July 15, 2013 and September 5, 2013). [SHL]

So also, *ḷ* > *ṭ* before a stop in euphonic combination in classical Tamil (see Rajam 1992: 106-107; Levitt 2010: 64-69). But before an enunciative vowel, *ḷ* is doubled (Rajam 1992: 107, 108-109).

In Vedic Sanskrit on the other hand, but consonant with this latter point, *ḍ* > *ḷ* when between two vowels. Thus, *īḍe* > *īḷe*, but *īḍya*. So also followed by *h*, for *ḍh*. Thus, *mīḍhuṣe* > *mīḷhuṣe*, but *mīḍhvān* (Whitney 1960[1889]: 19 [§54]). And so also in Prakrit (Pischel 1965[1957]: 172-173 [§240]). In North Indian Prakrit manuscripts, though, *l* is written, *ḷ* having vanished in both writing and spoken classical Sanskrit (Pischel 1965[1957]: 166-167 [§226]; see also Kuiper 1991: 71).

Pischel also notes that sometimes in Prakrit *t* and *d* > *l*; and as dentals are often represented by cerebrals in Prakrit (and sometimes vice versa, depending on dialect), through *ṭ* and *ḍ* they become *ḷ*. But as the North Indian manuscripts write *l* for *ḷ*, it cannot always be said with certainty whether *l* or *ḷ* is intended (1965[1957]: 174-175 [§244], 160-161 [§218]).

The Prakrits, of course, are commonly understood to reflect Dravidian speech.

Those correspondences that were indicated earlier by Illič-Svityč do not appear to be included by either Bomhard (2008) or Dolgopolsky (2008), as noted earlier (Section 1).[8] But given such alternations in both Afroasiatic and Dravidian, it is likely that we are going to find such in Nostratic in general.[9]

In Levitt (2010), a proposed relationship was spelled out in detail between Eng. *bubble*, *burble* and Lat. *bulla* 'bubble', both isolated within their IE sub-families, and Skt. *budbuda* 'bubble' and Te. *buḍabuḍa* 'with a bubbling noise' (*DEDR* 4249), which forms Levitt sees to be related on a genetic level (pp. 71-72).[10]

In Levitt (2003), a relationship similarly was suggested between Eng. *bone*, OHG. and ON. *bein* and Eng. *tube* < Lat. *tubus*, akin to Lat. *tuba* 'trumpet', both also of uncertain etymology and isolated within their respective IE sub-families – the *n* of *bone* etc. standing for *l* as noted above, and the etyma in *DEDR* 5050 Ka. *mūḷe* 'bone' and *DEDR* 4528 Te. *bomika* 'bone' (Pa. *būla*, Go. *būlā*, *bula*) (pp. 178-179).

In that place, there also was suggested a genetic relationship with competing forms for 'honey' given by Pokorny (2005[1959-69]/1: 707, 723-724), PIE **médhu* and PIE **meli-t* (Gen. **mel-nés*), with the etyma in *DEDR* 3268(a) Ta. *tīm* and (b) Ta. *tēṉ* 'sweet, honey, bee' (p. 178; see Section 2.6).

We can also note here a suggested relationship between the locative suffix *-il* in Tamil, which is related to words for 'place, house, home' in *DEDR* 494 Ta. *il* (Fabricius 1933: 88b), with the demonstrative pronoun in IE, PIE **to-*, **tā*, **ti̯o-* (Pokorny 2005[1959-69]/1: 1086-1087). Again, metathesis would be in force.

---

[8] See n. 3 above.

[9] Interestingly, note Lakota (= Sioux): North Dakota, South Dakota. In northern Lakota, or Dakota dialects in Alberta, Canada, the form is Nak(h)ota, for which point I must thank John Bengtson.

[10] All these forms are taken to be independent onomatopoetic developments, there being argument as to whether or not the Sanskrit word is a loan form from Dravidian or genetically inherited within IE. There is nothing very obvious about a very many onomatopoetic forms, though; and these are a case in point.

2.4. *Metathesis*

Metathesis has been referred to several times as being a formative process in Dravidian and Nostratic, and several examples have been given above. Neither Bomhard (2008, 2011) nor Dolgopolsky (2008) consider it as a formative process in Nostratic in their etymologies.

In Section 2.3 there was mentioned, among others, Levitt's earlier suggestion that PDr **kāl* 'leg, foot' (*DEDR* 1479) was cognate with Eng. *leg* etc. Without metathesis, in its meaning 'column, post' (Fabricius 1933: 235a; *DEDR* 1479), it can also be seen as being related to Eng. *column* < MF. *colomne* < Lat. *columna* (Devaneyan 1966: 236, 2004: 44-45). We can also suggest here a genetic relationship with Eng. *calf* (of the leg), ON. *kálfi*, "of unknown origin" (compare also, Ir., Gael. *calpa* 'leg, calf of the leg', *OED* 2: 781c; see also Devaneyan 2004: 45 [mng. 7], 'part of the leg below the knee').

Levitt (2000) dismissed K. C. A. Gnana Giri Nadar's connection of Ta. *nōkku* 'to look at, consider, reflect; n. sight, look, knowledge' (*DEDR* 3794) with Eng. *know*, OE. *cnāwen*, Lat. *gnoscere*, Levitt opting instead for Devaneyan's connection with Ta. *kāṇ* 'to see, behold, perceive, understand', *kāṇkai* 'knowledge' (*DEDR* 1443) and connecting Ta. *nōkku* instead with Eng. *look*, OE. *lōcian*, as had Devaneyan (pp. 413, 420-421, 424, 425). In the context of the clear-cut example of metathesis with regard to Eng. *leg*, though, and in the context of almost 60 sets of metathetical correspondences within Dravidian in Levitt (2003), many of the sets of correspondences containing more than one example, we ought now accept Gnana Giri Nadar's connection as also a correspondence. This would be still another example of metathesis between Dravidian and IE forms, though metathesis most likely took place within pre-Dravidian, and the IE forms for 'know' are probably related more directly to the etyma in *DEDR* 1443 Ta. *kāṇ*, *kāṇkai* etc.

We can also add here Ta. Ma. *kuḷam* 'tank, reservoir, lake', Tu. *kuḷa* 'tank, pond' (*DEDR* 1828) and Eng. *lake* < OF. *lac* < Lat. *lacus*, OE. *lagu* 'sea', Gk. *lakkos* 'pond' (*Webster's* 644a; *OED* 8: 595c-596a)

This ought to suffice to indicate in brief the validity of the thesis.

2.5. *Inserted and dropped r*

It has long been established that the name "Dravidian" [Skt. *draviḍa*] is a reflex in Sanskrit of "Tamil" [Ta. *tamiṟ*; by the standard transcription used for Tamil itself, *tamiḻ*]. Suggesting this connection are the Sanskrit variants *dramiḍa* and *dramila* and the Pali and Prakrit reflexes *damiḷa*, *daviḷa*, *daviḍa* (Caldwell 2000[1913]: 8-10; *CDIAL* 378b [no. 6632]).

Levitt (1998: 142-143 [no. 11]) suggested that Germ. *Deutsch* 'people, nation', and the common language of same, was also a reflex in IE of these names, with reverse *vṛddhi*-strengthening (*au* [*av*] → *u*) and metathesis of the vowels.[11]

So also, Levitt argued, the name "Druid" for the ancient inhabitants of Gaul and the British Isles was similarly related. The present etymology that relates the name to ancient magicians, sorcerers, and soothsayers and on an analogy with Irish, analyzes the form as *dru-vid* 'very knowing', would seem to read into the name later attributions, and be forced.

---

[11] Note that Caldwell thinks that the name "D(r)aviḍa" was the original, and "Tamil" a development.

There have been many suggestions as to the meaning of the name “Tamil”, but a meaning ‘people, nation’, used as well to refer to their common language, is supported from a literary aspect by Sivaraja Pillai (1936: 344-346; see Levitt 1998 write-up and Levitt 2009: 145, 2010: 22). Such is supported by the names of such tribal groups in the hills around the Assam Valley in northeast India, such as the Dafla and the Mizo, meaning ‘people’. See also in this regard Bailey (1959: 109-110), who notes that such ethnic names as “Goth”, the name “Evenki” for the Tungus, and the name “Bantu” derive from words meaning ‘man’ or ‘men’. So also, Bailey argues, the people called by the ethnic name Ir. *daha-*, OPers. *dahā*. And as well, the name of the Finno-Ugric “Mari” is said to have been borrowed from Indo-Iranian (Anonymous 2009; see Skt. *marya* ‘man, mortal; pl. people’ < PIE **meri̯o* [*MW* 791c; Pokorny 2005(1959-69)/1: 738-739]). See also Dolgopolsky (2008: no. 66), who attributes a meaning ‘member of one’s own ethnic group’ (→ ‘freeman’) for the PIE reflex of Skt. *ārya*, which carries a meaning in Sanskrit of ‘a respectable, honorable man, a man highly esteemed, noble’ (*MW* 152b; see Pokorny 2005[1959-69]/1: 67).

Levitt (1998) gave as parallel examples of matching forms both with and without *r* a suggested relationship between Eng. *fuck* and Germ. *pfrecken* ‘to prick’ (the latter word listed at Pokorny 2005[1959-69]/1: 167, the origin of the English word generally considered to be obscure), to which we can add colloq. Eng. *prick* ‘penis’; and between Eng. *speak*, *speech* and Germ. *sprechen* ‘to speak’ (Pokorny 2005[1959-69]/1: 997-998). Buck (1949: 18.21[5]) notes of the *r*-less forms in this latter set of etyma that their history is obscure.[12]

We can give here as well a number of other parallel examples:

(1) Etruscan: Tuscany. (2) Makran: Magan – the association of the Ancient Near Eastern place name “Magan” with the Makran coast has been suggested and supported by several people (Leemans 1960: 29; Hansman 1973: 554, 555, 557-558). Possehl (1996: 136), though, questions the etymological connection (which is here given support by parallel examples). (3) *arse* (obs., in polite use): *ass* (chiefly U.S.) (*OED* 1: 654c, 699a). [(4) *coss/cos*, *kos* < H. *kos*: Skt. *krośa* (*OED* 4: 65b, 2: 1012b).[13]] (5) *crow*: *caw* (*OED* 4: 65b, 2: 1012b) – also possibly related, *cry* (see Levitt 2008: 31b for a connection with Dravidian and Austroasiatic forms; so also, we can add in the context of the present paper, Eng. *scritch*, *screech*, considered by *OED* 14: 743c, 721a to be respectively “onomatopoetic” and an “echoic modification” of same). (6) *curse* (of unknown origin): *cuss* (U.S. colloq. or slang; in origin, a vulgar pronunciation or attenuation of *curse*) (*OED* 4: 151c, 4: 165c). (7) *drum* (Germanic; all the continental cognates have *tr-*): *tom-tom*, *tympanum* (*OED* 4: 1082a, 18: 215c, 784b). (8) *pork* (Lat. *porcus*): *pig* (etym. obscure) (*OED* 12: 134a, 11: 805b). (9) *tremor*: *temblor* (SWn U.S.) (*OED* 18: 470b, 17: 743a; Sánchez 2001: 2021a, 1952ab; see also Ernout and Meillet 1985: 700b, “La racine **ter-* ‘trembler’, qui a un charactère expressif, …”). (10) *trombone*, *trumpet*: *tuba*, *tube* (*OED* 18: 578a, 612b, 610bc, 630b, 640b; see Section 2.3 and Levitt 2003: 178-179 re: Eng. *bone*).

---

[12] With regard to Eng. *fuck*, see also the colloquial *friggin’* and *fricken’* as more socially acceptable forms of the word *fucking*.

[13] I bracket this example as dropping an *r* in a conjunct is normal development in words going from Old Indo-Aryan into Middle and New Indo-Aryan.

Many seemingly possible correspondences have been eliminated here on the basis of standing etymologies in *OED*, but they may in fact constitute reasonable competing etymologies, such as *crackle* (in some of its meanings) and *cackle*. Also note Fr. *clé*, *clef*, Lat. *clāuis*, *clāvus*: Eng. *key* (OFris. *kei*, *kay* – not in other Teutonic languages; ulterior etymology unknown) (*OED* 8: 404b).

In the examples given above, time and again it appears that forms with *r* are more polite, forms without the *r* are more vulgar or colloquial. This does not seem to apply in all cases, though.

Kuiper (1991: 70-79, 79-85) has written about an "intrusive" *r* and *ṛ* in Sanskrit loan forms. Often, the purpose of this is to "Sanskritize" foreign forms, but "often … the rationale of the insertion is far from clear" (p. 70). In the present context, we can see it as a general phenomenon in IE.

It has been said that such forms as Ta. *eṭṭi* 'merchant', Ta. *ēṇi* 'ladder', Ta. *cāvakar* 'ascetics', Ta. *cuṇṇam* 'powder', Ta. *paṭimam* 'idol', Ta. *vaṇṇam* 'color' come from, respectively, Skt. *śreṣṭhin* 'a distinguished man', Skt. *śreṇi* 'line, or row', Skt. *śrāvaka* 'disciple', Skt. *cūrṇa* 'powder', Skt. *pratimā* 'idol', and Skt. *varṇa* 'color' through Prakrit, which simplifies Sanskrit conjuncts (Vaidyanathan 1971). The observations here, though, throw such derivations in doubt. Such forms may, rather, be in some cases genetically related. Note that *TED* gives Tamil derivations for all these forms, whereas for such derivations as Ta. *nittam* 'dancing' < Skt. *nṛtta* 'id.' and Ta. *aramiyam* 'palace, terrace of a house' < Skt. *harmya* 'palace, large house' given by Vaidyanathan, *TED* concurs.

2.6. *Miscellaneous*

Many of the forms preserved in Dravidian prove to be very ancient from the vantage of IE, and even from the vantage of Afroasiatic. This demonstrates the antiquity of forms preserved in Dravidian within Nostratic.

Etyma as in *DEDR* 5086 Ta. *mē … mēl*, *mēlimai*, *mēlukka*, *mēlai*, *mēlōr* etc. refer in the main to 'excellence, height, superiority, eminence'. By means of a syncope within Proto-Dravidian, these forms are related to the etyma in *DEDR* 4841 Ta. *micai*, which forms have similar cognitions (Krishnamurti 2003: 96). It has been argued elsewhere that these forms are related on a Nostratic level to Heb. *melekh* 'king', Ar. *malik* 'king, sovereign, monarch', Ar. *malaka i (malk, milk, milk)* 'to take in possession … to rule, reign, exercise power or authority', Ar. *mālik* 'ruling, owning; owner, proprietor, master' etc., the Afroasiatic root showing a common Proto-Dravidian suffix added to a root shown in its more basic form in Dravidian (Levitt 2009: 136-140, 148-150).

Etyma in *DEDR* 3268(a) Ta. *tīm* and (b) Ta. *tēṉ* contain words for 'sweet, honey, bee'. PIE **médhu*, Skt. *madhu* and PIE **meli-t* 'honey' appear to be formed by metathesis of a Dravidian form with a common Proto-Dravidian suffix added (Levitt 2003: 178). In that the initial *m-* of the IE forms corresponds to a suffix in the Dravidian forms speaks to the priority of the Dravidian forms.

The etyma in *DEDR* 3266 Ta. *tī* carry such meanings as 'fire, lamp, heat, light, splendor, shine'. These have a metathetic reflex within Dravidian of the etyma in *DEDR* 829 Ta. *el* which carry such meanings as 'lustre, splendor, light, sun, daytime, shine, spark'. The etyma in *DEDR* 3266 Ta. *tī* appear to be cognate within IE to such forms as Skt. *div*, *dyu* 'heaven, sky, day' (Gk. *Zeus*), giving rise to the root *dyut-* 'to shine, be bright or brilliant', Skt. *deva* (Lat. *dīvus*, *deus*) 'a deity, god; heavenly, divine', Skt. *divya* 'divine, heavenly' etc. (Pokorny 2005[1959-69]/1: 183-187; Devaneyan 2004: 91-95; Bomhard 2011: 79-80 [no. 149]; see Section 2.3).[14]

Notice Eng. *stand*, *stall* < PIE **stā*: **stə*, Skt. *sthā*. Levitt (1998) observed that IE forms appear to show common Proto-Dravidian suffixes in a number of instances, here *-nt* and *-l*. So also, notice *DEDR* 4345 Ta. *pū* 'flower, blossom, bloom, flourish': Skt. *phulla* 'a full-blown flower, blown (as a flower), abounding in flowers, flowery', which speaks to the antiquity of the Dravidian forms as they appear without the suffix *-l*. The Sanskrit form is without clear-cut motivation in IE (see Levitt 2000: 429-430 [no. 4]).

Also, note Ta. *taka-tak-eṉal* 'onomatopoetic expression of dazzling, glowing, glittering'; Te. *daga-daga* 'glitter', Kur. *dagnā* 'to light, set fire to, burn', *dagmā* 'to catch fire, be burnt' (DEDR 2998): Skt. *dah-* 'to burn, scorch, consume by fire', *dagdha* 'burnt, scorched, consumed by fire' (PIE **dheg*$^{\text{u̯}}$*h*, Pokorny 2005[1959-69]/1: 240-241; Bomhard 2011: 61 [no. 90])[15]; Ta. *tī* … (*-pp-*, *-tt-*) … *tīppi*, *tīppu* 'to be burnt, charred … to shine; n. fire, lamp, heat, light, splendor' (*DEDR* 3266; see above and Section 2.3): Skt. *dīp-* 'to blaze, flare, shine, be luminous', *dīpa* 'light, lamp, lantern (Pokorny 2005[1959-69]/1: 183; *MW* 481a).

To change focus for a moment, in the context of Vasileiadis (2007) regarding mythological connections between ancient Greece and the eastern Mediterranean, and the Coorgs in Karnataka in western India, I could not help but recall the connection made by Legrand (1954: 44) between ancient Greek *gala* 'milk' and the etyma in *DEDR* 4096 Ta. *pāl* 'id.' Legrand draws attention to a Kanarese form *hālu* for 'milk', not listed in *DEDR* but it is well-known that *p* > *h* in Kanarese, prevalent first in Kanarese prose writing of the 10th c. CE, and shared as well with Tuḷu, Baḍaga, and Kuṟumba dialects (Zvelebil 1970: 86; compare Krishnamurti 2003: 120-121). Legrand notes that Greek *g* is the same as *h* in many words.

Webster's 470a (s.v. *galaxy*) notes Gk. *gala* to be akin to Lat. *lac*, *lactis* 'milk' (< **glact*, Pokorny 2005[1959-69]/1: 400-401).

The Greek form, though, may speak to a loanword connection between Dravidian and Greek, and to this areal Dravidian sound change having taken place earlier than assumed at present. See, for instance, a 1989 paper presented by Dr. Uppangala Rama Bhat at the 17th All India Conference of Dravidian Linguists and printed in Bhat (2006) which points out that *p* > *h* in Kanarese is already in evidence in the earliest Kanarese inscriptions dating from the mid-4th c. to mid-5th c. CE.

Such loanword possibility ought not be overlooked when etymologizing certain other Greek forms as well, such as Gk. *dendron* 'tree'. This Greek form is often given as an IE justification for Skt. *daṇḍa* 'stick, staff, punishment, army' – which Sanskrit form Levitt has argued, to the contrary, is rather a loan form from Dravidian in Sanskrit, as the removal of its Dravidian reflexes from the Dravidian lexicon leaves a gaping hole in an otherwise logical semantic spread for Dravidian forms of the same and similar phonological shape (Levitt 2003: 183-184 [nos. 28, 29]; *KEWA* 2: 11-12, *EWA* 1: 691-692; see also Pokorny 2005[1959-69]/1: 194, 214).

---

[14] Dolgopolsky (2008), however, lists the IE forms with his no. 2241, and the Dravidian forms under his no. 2250.

[15] See also Dolgopolsky (2008: no. 511), which does not correlate the IE forms with Dravidian forms, however, as does Bomhard (2011).

3. Summation

Dravidian has been genetically related by specialists in different language families to languages all over the world, which efforts have generally received mixed reviews. It is hoped that by focusing here on Dravidian and IE, some of the reasons for these connections may emerge, and that a clearer idea of the development of Nostratic languages and of Nostratic morphology may emerge.

Abbreviations

CDIAL = Turner, Ralph Lilley. 1966-1985. *A Comparative Dictionary of Indo-Aryan Languages.* London: Oxford University Press, 1966. *Indexes*, compiled by Dorothy Rivers Turner, 1969. *Phonetic Analysis*, by Ralph Lilley Turner & Dorothy Rivers Turner, 1971. *Addenda and Corrigenda*, edited by J. C. Wright, 1985.

DBIA = Burrow, Thomas & Murray Barnson Emeneau. 1962. *Dravidian Borrowings from Indo-Aryan.* University of California Publications in Linguistics, 26. Berkeley & Los Angeles: University of California Press.

DEDR = Burrow, Thomas & Murray Barnson Emeneau. 1984. *A Dravidian Etymological Dictionary.* 2nd edition. Oxford: Clarendon Press.

EWA = Mayrhofer, Manfred. 1992-2001. *Etymologisches Wörterbuch des Altindoarischen*, 3 volumes. Heidelberg: Carl Winter, Universitäts-Verlag.

KEWA = Mayrhofer, Manfred. 1956-1980. *Kurzgefasstes etymologisches Wörterbuch des Altindischen*, 4 volumes. Heidelberg: Carl Winter, Universitätsverlag.

MW = Monier-Williams, Monier. 2000[1899]. *A Sanskrit-English Dictionary, Etymologically and Philologically Arranged.* New edition. Oxford: The Clarendon Press.

OED = 1989. *The Oxford English Dictionary.* 2nd edition. Prepared by J. A. Simpson & E. S. C. Weiner. 20 volumes. Oxford: Clarendon Press.

TED = Devaneyan, Gnanamuthu [Tēvanēyan̲, Ñānamuttan̲]. 1985-2011. *Centamil̲c cor̲pir̲appiyal pērakaramutali* [A Comprehensive Etymological Dictionary of the Tamil Language; Editors vary]. 12 volumes (in 31 volumes). Chennai: Directorate of Tamil Etymological Dictionary Project.

Webster's = 1974[1973]. *Webster's New Collegiate Dictionary.* Springfield, Massachusetts: G. & C. Merriam Company.

Bibliography

[Anonymous.] 2009. "Mari Language". *Wikipedia, The Free Encyclopedia* [website]. http://en.wikipedia.org/wiki/Mari_language (accessed 11 March 2009).

Bailey, Harold Walker. 1959. "Iranian *arya-* and *daha-*". *Transactions of the Philological Society* 58.1: 71-115.

Bhat, Uppangala Rama. 2006. "*p > h*, A Phonetic Change in Kannada". In Uppangala Rama Bhat, *A Chat with Scholars (A Collection of Research Papers)*. Udupi, Karnataka State: Akalanka Prathisthana (R) Udupi, 18-24.

Blažek, Václav. 1992. "Australian Elements in Dravidian Lexicon?" In Vitaliĭ Viktorovich Shevoroshkin (Ed.), *Nostratic, Dene-Caucasian, Austric and Amerind: Materials from the First International Interdisciplinary Symposium on Language and Prehistory, Ann Arbor, 8-12 November, 1988*. Bochum: Universitätsverlag Dr. Norbert Brockmeyer, 421-431.

Blažek, Václav. 2006. "Was There Australian Substratum in Dravidian?" *Mother Tongue* 11: 275-294.

Bomhard, Allan R. 2008. *Reconstructing Proto-Nostratic: Comparative Phonology, Morphology, and Vocabulary*. Leiden: E. Brill. http://www.columbia.edu/cgi-bin/cul/resolve?clio9183101 (= http://iedo.brillonline.nl//dictionaries/content/nostratic/index.html) (accessed 10 January 2012 and 13 February 2012).

Bomhard, Allan R. 2011. *The Nostratic Hypothesis in 2011, Trends and Issues*. Journal of Indo-European Studies Monograph Series, 59. Washington, D.C.: Institute for the Study of Man.

Brown, George William. 1930. "The Possibility of a Connection between Mitanni and the Dravidian Languages". *Journal of the American Oriental Society* 50.4: 273-305.

Buck, Carl Darling. 1949. *A Dictionary of Selected Synonyms in the Principal Indo-European Languages*. Chicago: University of Chicago Press.

Burrow, Thomas. 1945. "Some Dravidian Words in Sanskrit". *Transactions of the Philological Society* 44.1: 79-120.

Burrow, Thomas. 1947. "Dravidian Studies VI: The Loss of Initial *c/s* in South Dravidian". *Bulletin of the School of Oriental and African Studies, University of London* 12.1: 132-147.

Burrow, Thomas. 1955. *The Sanskrit Language*. London: Faber & Faber.

Burrow, Thomas. 1973. *The Sanskrit Language*. New and revised edition. London: Faber & Faber.

Caldwell, Robert. 2000[1913]. *A Comparative Grammar of the Dravidian or South-Indian Family of Languages*. 3rd edition. Revised and edited by J. L. Wyatt and T. Ramakrishna Pillai. Chennai: University of Madras.

Cohen, David. 2010. *Dictionnaire des racines sémitique ou attestées dans les langues sémitique, comprenant un ficher comparatif de Jean Cantineau*, fasc. 9. Paris: Peeters.

David, H. S. 1966. "Suggestions to Research Schools and Lexicographers in Tamil and Dravidology". *Tamil Culture* 12.1: 5-12, 12.2/3: 231-246, 12.4: 269-285.

Devaneyan, Gnanamuthu [Tēvanēyan̠, Ñānamuttan̠]. 1966. *The Primary Classical Language of the World*. Ka[t]padi Extension, North Arcot Dt., Madras State: Nesamani Publishing House.

Devaneyan, Gnanamuthu. 2004. *Nostratics – The Light from Tamil According to Devaneyan (1977-80 Studies of G. Devaneyan on the Spread in Different Language Families of the World of Derivatives from 22 Basic Tamil Words)*, translated by P. Ramanathan. Chennai: The Tirunelveli South India Saiva Siddhanta Works Publishing Society, Ltd.

Dixon, Robert M. W. 1980. *The Languages of Australia*. Cambridge: Cambridge University Press.

Dixon, Robert M. W. 2002. *Australian Languages, Their Nature and Development*. Cambridge: Cambridge University Press.

Dolgopolsky, Aharon. 2008. *Nostratic Dictionary* [published online]. Cambridge: The McDonald Institute. http://www.dspace.cam.ac.uk/handle/1810/196512 (accessed several times in March, April, and July 2012).

Emeneau, Murray Barnson. 1967[1957]. "Numerals in Comparative Linguistics (with Special Reference to Dravidian)". In Murray Barnson Emeneau, *Dravidian Linguistics, Ethnology and Folktales: Collected Papers*. Annamalai University Department of Linguistics Publication, 8. Annamalainagar: Annamalai University, 139-147.

Emeneau, Murray Barnson. 1988. "Proto-Dravidian *c- and Its Developments". *Journal of the American Oriental Society* 108.2: 239-268.

Ernout, Alfred & Antoine Meillet. 1985. *Dictionnaire étymologique de la langue latine: Histoire des mots*, 4. édition, 4. tirage, augmenté d'additions et de corrections nouvelles, par Jacques André. Paris: Éditions Klincksieck.

Fabricius, Johann Philip. 1933. *A Dictionary, Tamil and English, Based on Johann Philip Fabricius's "Malabar-English Dictionary"*. 3rd edition, revised and enlarged. Tranquebar: Evangelical Lutheran Mission Publishing House.

Fähnrich, Heinz. 1965. "Iberokaukasisch und Dravidisch". *Bedi Kartlisa, revue de kartvélologie* 19-20: 136-158.

Fähnrich, Heinz. 1981. "Das Sumerische und die Kartwelsprachen". *Georgica* 4: 89-101.

Fähnrich, Heinz. 1991. "Kartwelisch-dravidische Sprachparallelen". *Zeitschrift für Phonetic, Sprachwissenschaft und Kommunikationsforschung* 44: 339-343.

Field, Julie S. & Marta Mirazón Lahr. 2005. "Assessment of the Southern Dispersal: GIS-Based Analyses of Potential Routes at Oxygen Isotopic Stage 4". *Journal of World Prehistory* 19.1: 1-45.

Goodenough, Ward H. 1970. "The Evolution of Pastoralism and Indo-European Origins". In George Cardona, Henry M. Hoenigswald, & Alfred Senn (Eds.), *Indo-European and Indo-Europeans, Papers Presented at the Third Indo-European Conference at the University of Pennsylvania*. Philadelphia: University of Pennsylvania Press, 253-265.

Haarmann, Harald. 1998. "On the Problem of Primary and Secondary Diffusion of Indo-Europeans and Their Languages". *The Journal of Indo-European Studies* 26.3/4: 391-419.

Hansman, John. 1973. "A *Periplus* of Magan and Meluḫḫa". *Bulletin of the School of Oriental and African Studies, University of London* 36.3: 554-587.

Hodge, Carlton T. 1998. "The Implications of Lislakh for Nostratic". In Joseph C. Salmons & Brian D. Joseph (Eds.), *Nostratic, Sifting the Evidence*. Amsterdam Studies in the Theory and History of Linguistic Science, 142. Series IV: Current Issues in Linguistic Theory. Amsterdam & Philadelphia: John Benjamins Publishing Company, 237-256.

Illič-Svityč, Vladislav M. 1989. "The Early Reconstructions of Nostratic by V. M. Illič-Svityč", translated and arranged by Mark Kaiser. In Vitaliĭ Viktorovich Shevoroshkin (Ed.), *Reconstructing Languages and Cultures: Abstracts and Materials from the First International Interdisciplinary Symposium on Language and Prehistory, Ann Arbor, 8-12 November, 1988*. Bochum: Studienverlag Dr. Norbert Brockmeyer, 125-176.

Illič-Svityč, Vladislav M. 1990. "The Nostratic Reconstructions of V. Illich-Svitych", translated and arranged by Mark Kaiser. In Vitaliĭ Viktorovich Shevoroshkin (Ed.), *Proto-Languages and Proto-Cultures: Materials from the First International Interdisciplinary Symposium on Language and Prehistory, Ann Arbor, 8-12 November, 1988*. Bochum: Universitätsverlag Dr. Norbert Brockmeyer, 138-167.

Kivisild, Toomas et al. 2003. "The Genetic Heritage of the Earliest Settlers Persists Both in Indian Tribal and Caste Populations". *The American Journal of Human Genetics* 72: 313-332.

Klimov, Georgiĭ Andreevich. 1998. *Etymological Dictionary of the Kartvelian Languages*. Trends in Linguistics; Documentation, 16. Berlin & New York: Mouton de Gruyter.

Konow, Sten. 1903. "Notes on Dravidian Philology". *The Indian Antiquary* 32: 449-458.

Krishnamurti, Bhadriraju. 1961. *Telugu Verbal Bases, A Comparative and Descriptive Study*. University of California Publications in Linguistics, 24. Berkeley & Los Angeles: University of California Press.

Krishnamurti, Bhadriraju. 2003. *The Dravidian Languages*. Cambridge: Cambridge University Press.

Kuiper, Franciscus Bernardus Jacobus. 1991. *Aryans in the Rigveda*. Leiden Studies in Indo-European, 1. Amsterdam and Atlanta, Georgia: Rodopi.

Kuznetsov, P. 2005. "An Indo-Iranian Symbol of Power in the Earliest Steppe Kurgans". *The Journal of Indo-European Studies* 33.3/4: 325-338.

Leemans, Wilhelmus François. 1960. "The Trade Relations of Babylonia and the Question of Relations with Egypt in the Old Babylonian Period". *Journal of the Economic and Social History of the Orient* 3.1: 21-37.

Legrand, F. 1954. "Tamil Loan Words in Greek". *Tamil Culture* 3.1: 36-45.

Levitt, Stephan Hillyer. 1989. "The Alternation of *r* and *l* in Dravidian". *Indian Linguistics* 50: 130-146. Errata in *Indian Linguistics* 52 [1991]: 149-150.

Levitt, Stephan Hillyer. 1998. "Is There a Genetic Relationship Between Indo-European and Dravidian?" *The Journal of Indo-European Studies* 26.1/2: 131-159

Levitt, Stephan Hillyer. 2000. "Some More Possible Relationships Between Indo-European and Dravidian". *The Journal of Indo-European Studies* 28.3/4: 407-438.

Levitt, Stephan Hillyer. 2003. "The Alternation of **-l-/*-ḷ-* and **t-* in Metathetical Forms in Dravidian". *Folia Linguistica Historica* 24.1/2: 175-192.

Levitt, Stephan Hillyer. 2007a. "A Word for 'Horse' in Chinese and Dravidian". *International Journal of Dravidian Linguistics* 36.2: 19-44.

Levitt, Stephan Hillyer. 2007b. "Dravidian and Indo-European". *Chemmozhi, A Quarterly Newsletter from Centre of Excellence for Classical Tamil* 2.1: 20-22.

Levitt, Stephan Hillyer. 2008. "Dravidian and Austroasiatic". *Chemmozhi, A Quarterly Newsletter from Central Institute of Classical Tamil* 2.2/3: 31-33.

Levitt, Stephan Hillyer. 2009. "The Ancient Mesopotamian Place Name 'Meluḫḫa'". *Studia Orientalia* 107: 135-176.

Levitt, Stephan Hillyer. 2010. "Fortunatov's Law and Dravidian". *International Journal of Dravidian Linguistics* 39.1: 19-90.

Levitt, Stephan Hillyer. 2011-12. "The Alternation of Initial Consonants in Dravidian". *International Journal of Dravidian Linguistics* 40.2: 129-146 and 41.1: 38-97.

Levitt, Stephan Hillyer. 2012. "The Number 'One'", together with comments from four discussants and a 'Response to Discussants'". *Mother Tongue* 17: 103-166 [paper proper, 103-118; response, 149-166].

Masica, Colin P. 2001. "The Definition and Significance of Linguistic Areas: Methods, Pitfalls, and Possibilities (with Special Reference to the Validity of South Asia as a Linguistic Area). In Peri Bhaskararao and K. V. Subbarao (Eds.), *The Yearbook of South Asian Languages and Linguistics, 2001*. London: SAGE, 205-267.

McAlpin, David. 1981. *Proto-Elamo-Dravidian: The Evidence and Its Implications.* Transactions of the American Philosophical Society, 71.3. Philadelphia: The American Philosophical Society.

Morris, Jonathan. 2006. "Trombetti, Gatti and the Birth of the Indo-Pacific Concept". *Mother Tongue* 11: 91-121.

National Geographic Society & IBM. 2005-2006. *The Genographic Project, A Landmark Study of the Human Journey* [A research partnership of National Geographic & IBM / Global field sciences supported by the Waitt Family Foundation], *Journey of Man*, narrated by Spencer Wells. [Washington, D.C.?]: National Geographic Society & IBM (DVD video with 19 page pamphlet).

Ohno, Susumu [Ōno, Susumu]. 1983. "The Loss of Initial *c* in Tamil and *s* in Japanese". *Uyarāyvu* [University of Madras] 1: 16-21.

Oppenheimer, Stephen. 2007[2004]. *The Real Eve, Modern Man's Journey out of Africa*. New York: Carroll & Graf Publishers.

Patel, Samir. 2007. "World Roundup". *Archeology* 60.5: 14-15.

Pei, Mario. 1962. *The Families of Words*. New York: St. Martin's Press.

Pischel, Richard. 1965[1957]. *Comparative Grammar of the Prākṛit Languages*, translated by Subhadra Jhā. Delhi: Motilal Banarsidass. [Reprint adds index by Martino De Zilva Wickremasinghe, published earlier in parts as supplements to *Indian Antiquary* 34-38 (1905-1909).]

Pokorny, Julius. 1960. "Keltische etymologien". *Indogermanische Forschungen* 65: 161-171.

Pokorny, Julius. 2005[1959-69]. *Indogermanisches etymologisches Wörterbuch*, 2 volumes. Tübingen & Basel: A. Francke Verlag.

Possehl, Gregory L. 1996. "Meluhha". In Julian Reade (Ed.), *The Indian Ocean in Antiquity*. London & New York: Kegan Paul International in association with The British Museum, 133-208.

Rajam, V. S. 1992. *A Reference Grammar of Classical Tamil Poetry (150 B.C. – Pre-Fifth/Sixth Century A.D.)*. Memoirs of the American Philosophical Society, 199. Philadelphia: American Philosophical Society.

Ramanathan, P. [Irāmanātaṉ, Pi.] 1983-84. "Āstirēliyap paḻaṅkuṭi makkaḷ moḻikaḷum tamiḻum" [The relationship between the Australian aboriginal languages and Tamil]. *Centamiḻc Celvi* 58: 359-362.

Ramanathan, P. 2009a. Tamiḻum āttirēliyap paḻaṅkuṭi makkaḷ moḻikaḷum [The relationship between Tamil and the Australian aboriginal languages]. *Chemmozhi, Ceyti Maṭal* [Newsletter] (in Tamil) 4: 40-42.

Ramanathan, P. 2009b. "Affinities of Tamil with 'Pre-Indo-European' Languages". C*hemmozhi: A Quarterly Newsletter from Central Institute of Classical Tamil* 3.2/3: 35-37.

Ramaswami Aiyar, L. V. 1929a. "Notes on Dravidian II (iv-vi)". *The Indian Historical Quarterly* 5.1: 145-153.

Ramaswami Aiyar, L. V. 1929b. "The Linguistic History of Certain Dravidian Words". *Journal of the Department of Letters, University of Calcutta* 19.8: 1-14.

Ramaswami Aiyar, L. V. 1930. "Kui Words and Dravidian Origin". *Journal of Oriental Research, Madras* 4.2: 166-180.

Ramaswami Aiyar, L. V. 1932. "The Interrogative Bases of Dravidian". *The Indian Antiquary* 61: 3-5, 25-30.

Ramaswami Aiyar, L. V. 1935-38. "Dravidic Sandhi". *The Quarterly Journal of the Mythic Society* n.s. 25.4: 265-270, 26.1/2: 91-112, 26.3/4: 169-185, 27.1/2: 35-53, 27.3/4: 239-265, 28.1: 20-27, 28.2: 94-110, 28.4: 306-311.

Sahoo, Sanghamitra et al. 2006. "A Prehistory of Indian Y-Chromosomes: Evaluating Demic Diffusion Scenarios". *Proceedings of the National Academy of Sciences of the United States of America* 103.4: 843-848. http://www.pnas.org (accessed 7 August 2007).

Sánchez, Aquilino. 2001. *Gran diccionario de uso del español actual*. Alcobendas, Madrid: Sociedad General Española de Librería.

Sayers, William. 2008. "The Etymologies of English *dog* and *cur*". *The Journal of Indo-European Studies* 36.3/4: 401-410.

Sivaraja Pillai, K. Narayan. 1936. "The Derivation of the Word 'Tamil' – A Philological Causerie". In *Dr. S. Krishnaswami Aiyangar Commemoration Volume*. Madras: The Committee, 339-347.

Szalek, Benon Zbigniew. 2006. *Lycian, Lydian and Other Ancient Languages in the Light of Heuristics and Cryptology: (A Research Report)*. Szczecin: B. Z. Szalek.

Telegin, D. Ya. 2005. "The Yamna Culture and the Indo-European Homeland Problem". *The Journal of Indo-European Studies* 33.3/4: 339-357.

Trombetti, Alfredo. 1906-20. "Saggi di glottologia generale comparata. I. I pronomi personali: Introduzione; Parte prima". *Memorie della R. Academia delle Scienze dell'Istituto di Bologna, Classe de Scienze Morali, Sezione di Scienze Storico-Filologiche*, s. 1, t. 1: 5, 12-138, 161-263. "Parte seconda", s. 1, t. 2: 35-160. "II. I numerali: Introduzione; Parte prima: I. Ottentoto-Boschimano, II. Bantu, III. Lingua delle zona centrale africana, IV. Camitosemitico, V. Caucasico, Basco e lingue affini, VI. Indoeuropeo, VII. Uralaltaico e lingue paleoasitiche", s. 1, t. 3: 10-181. "VIII. Dravidico-Australiano, IX. Munda-Polinesiaco, X. Indocinese, XI. Lingue dell'America", s. 1, t. 4: 3-171. "XI. Lingue dell' America (continuazione)", s. 1, t. 5: 51-85. "I numerali. Parte seconda. I. Il numerale 1, II. I numerali 2 e 4, III. Il numerale 3, IV. I numerali 6, 7, 8 e 9, V. I numerali di origine nominale, VI. Osservazioni sui numerali, VII. Deduzioni per gli aggruppamenti linguistici. Osservazioni e aggiunte", s. 1, t. 7: 3-94. "III. Comparazioni lessicali", s. 1, t. 9: 17-85; s. 1, t. 10: 3-71; s. 2, t. 1: 13-109; s. 2, t. 2: 3-99; s. 2, t. 3: 3-83; s. 2, t. 4: 3-61. Bologna: Tipografia Gamberini e Parmeggiani & Industrie Grafiche Italiane.

Vaidyanathan, S. 1971. *Indo-Aryan Loanwords in Old Tamil*. Madras: Rajan Publishers.

Vasileiadis, Antonios. 2007. "New Evidence of Earlier Contacts Between West India and East Mediterranean". *International Journal of Dravidian Linguistics* 36.2: 1-18.

Whitney, William Dwight. 1960[1889]. *Sanskrit Grammar*. 2nd edition. Cambridge, Massachusetts: Harvard University Press.

Wood, Michael. 2007. *The Story of India*. London: BBC Books [Random House].

Zvelebil, Kamil. 1970. *Comparative Dravidian Phonology*. Janua Linguarum; Series Practica, 80. The Hague: Mouton.

Zvelebil, Kamil. 1972. "The Descent of the Dravidians". *International Journal of Dravidian Linguistics* 1.2: 57-63.

MOTHER TONGUE
Journal of the Association for the Study of Language in Prehistory • **Issue XIX • 2014**

Burushaski and the Western Dene-Caucasian Language Family: Genetic and Cultural Linguistic Links

John D. Bengtson
Association for the Study of Language in Prehistory

This paper is divided into three major parts: (a) some of the morphological evidence for the genetic relationship of the Burushaski language to Yeniseian, North Caucasian, and Basque languages; (b) some of the lexical evidence (both basic and cultural) for the same; and (c) phonological correspondences.[1]

The hypothesis explored here, that the mysterious and 'isolated' Burushaski language of the high mountains of Pakistan is genetically related to Yeniseian, North Caucasian, and Basque (and more remotely to other Dene-Caucasian languages), has been refined from several earlier proposals.

Apart from a few fleeting observations, one of the earliest forerunners of the hypothesis was outlined by the Austrian Robert Bleichsteiner (1930), who compared Burushaski primarily with Caucasian, and explicitly also with the other "Japhetic" languages (Marr's term), such as Basque, Iberian, Etruscan and Sumerian. His grammatical comparisons were mainly typological, noting for example (p. 299) that the Dagestani (EC) languages Andi, Tsezi, Lak, Archi, Rutul, Tsakhur, Budukh possess more or less "the same" four nominal classes as Burushaski: (I) male human, (II) female human, (III) other animate beings and various things (inanimates), and (IV) all other inanimates (see further below). He also remarked on the multiplicity of plural suffixes in Burushic and EC (pp. 303-306), and offered a number of material correspondences, e.g., Burushaski *har* 'ox', pl. *har-**o*** ~ Bezhta *os* 'ox', pl. *os-**o***; Bur *ǰi* 'soul', pl. *ǰi-**ming*** ~ Archi *anš* 'apple', pl. *anš-**um***, etc. Bleichsteiner pointed out some typological similarities between Burushaski and EC case markers, but not many of the material comparisons can be confirmed.

Bleichsteiner's pronoun comparisons (pp. 310-311) were generally quite good, and some conform to present-day DC etymologies, such as Bur 1st pers. sing. *ža, ǰa, ǰe* ~ Udi *zu*, Lezgi *zun*, etc.; Bur 2nd pers. sing. *un, um, ung* ~ Udi *un*, Lezgi, Agul *wun*, etc.; Bur demonstratives *ke, ku, gu* ~ Botlikh *go-*, etc.; Bur *men* 'who?' ~ Chechen *mi*-la, etc. (Some of these comparisons included Kartvelian forms as well.)[2] Bleichsteiner (pp. 319-331) offered 73 lexical comparisons, but these have not fared so well in the light of present-day DC etymologies. Of these only a handful coincide with our current hypothesis, e.g.:[3]

---

[1] Based on a paper given at the 3rd Harvard Round Table on Ethnogenesis of South and Central Asia, Harvard University, May 12-14, 2001, with the title "Genetic and Cultural Linguistic Links between Burushaski, Caucasian Languages and Basque." It has been revised (2014) and updated according to the current model of the Dene-Caucasian hypothesis. See also the **Postscript** at the end of the article. I am thankful to Peter Rowley-Conwy for assistance with the dating of animal domestication, and to Michael Witzel, Václav Blažek, and George Starostin for their discussion, support and encouragement.

[2] Since the taxonomic distinction between Kartvelian ("South Caucasian") and North Caucasian was not well-established until some decades later, the inclusion of Kartvelian comparanda with Burushaski (and Basque) was also frequent in the works of Bouda, Lafon, Uhlenbeck, and others. Most paleolinguists now consider Kartvelian to be a subgroup of the Nostratic (Eurasiatic) macrofamily, or closely related to it (e.g. J.H. Greenberg).

> §19. Bur (Y) *ɣóqares,* (H) *ɣókuras*, (N) *ɣókurac* 'raven' ~ Avar *qaqra* 'raven' [? – not verified; but cf. Chechen *q'ig* 'crow', Lezgi *k'ʷaʁ* 'raven', Agul *q:ʷaR̀anaj* 'crow', etc. < PNC **q̇HwāɣV*. However, these words are onomatopoetic, cf. PIE **grāk-* 'rook, daw, crow'; Turkic **KArga* 'crow', etc.].
>
> §21. Bur (Y) *ɣórkun*, (H) *ɣúrqun*, (N) *ɣúrquc* 'frog' ~ Avar *q':ʷerq':*, Khinalug *q'urq'or*, etc. (see below) [onomatopoeic, but phonetic forms are quite close].
>
> §32. Bur (Y) *=húṭ-is*, (H, N) *=úṭ, =úṭ-is* 'foot' ~ Avar *ħet'é / ħet'* 'foot', etc. (see below).
>
> §44. Bur (H, N) *ḍunḍú* 'bee, beetle' ~ Tsakhur *t'ot'* 'bee', Archi *t'ant'*'bee', Dargwa *t'ent' ~ t'ant'*'fly', etc. (PNC **ṭə̆mṭō*).

Some others, such as §9a: Bur (Y) *gacér*, (H, N) *gachír* 'vulture' ~ Andi *qačirɣa* 'vulture' [Godoberi *qarč:iʁa*, Tsakhur *q'aǯir*, Rutul *ɢaǯir,* etc. (SKJa 227)] look interesting at first glance, but a closer examination raises doubts. Berger (1998 III: 141) refers the Bur word to Pali *gijjha-* 'greedy; vulture'; Waigali, Dameli *grı̄̃č* 'kite', etc. (CDIAL 4430), which seems phonetically improbable. The Caucasian words resemble Mongolian *qaǯir* 'vulture', and may be borrowed from Kalmyk, a Mongolic language in the Caucasus region, or a related source, and the Bur word probably has a similar origin. (The Mongol Empire formerly included the Burushic area.) In sum, these words for 'vulture' are not genetic DC cognates, but loanwords from Mongolic sources. Obviously, this kind of winnowing process needs to be performed on all "Dene-Caucasian" comparisons proposed in the past (and present).

In 1950 Karl Bouda's article "Die Sprache der Buruscho" appeared. On the first page Bouda expounded his view that Burushaski seemed to be a keystone that connected the great archaic language families (today we would say "macrofamilies"), including "Indo-Chinese", Yeniseian, Euskaro-Caucasian, even Austronesian, Austroasiatic, and Chukchi. The article divides lexical and grammatical comparisons into three parts: I. "Indo-Chinese" connections (including Yeniseian!), II. Caucasian connections (the great majority from [North] Caucasian; a few from Kartvelian), and III. Chukchi connections.

Of Bouda's 86 "Indo-Chinese" comparisons very few Bur + ST parallels can be verified, possibly only:

> §48 Bur *gi* [H *ɣiín*, N *ɣáin*, Y *ɣēn* "thief", *ɣé-* "to steal"] ~ Tibetan r-*ku* "to steal" [< PST *(r-)*qhō* – same etymology as Bouda (1954: §30), discussed below].
>
> §56 Bur *ghar* [*ɣar*] "song", *ghare* [*ɣar´-*] "speak" ~ Tibetan d-*gyer* "to sing" [< PST **k(j)ār* or **g(j)ār*].

Though we cannot exclude that further study might increase this total. Similarly, very few of Bouda's Burushaski-Yeniseian comparisons (mixed into the "Indo-Chinese" section) withstand the winnowing process and are retained in the current DC inventory:

---

[3] Updated and corrected transcriptions have been added. Note that many of the lexical comparisons cited in this initial section are more fully documented and analyzed in the main **Lexical Evidence** and **Phonological Correspondence** sections, later in this paper.

§28 Bur =*s*- ‘say’ ~ PY **ʔas*- / **ʔes*- ‘shout, call’ [cf. Bsq **e=śa-n* ‘to say’, **hic* ‘word’; Tsezi =*es*- ‘to say, speak’, etc.].
§38 Bur *śi* ‘eat’ ~ Ket *sī* id. [cf. Bsq **auśi*-ki ‘to bite’; Tsezi =*ats’*- to eat’, etc.].
§63 Bur =*ik* ‘name’ ~ Ket *ī*, Kott *ix, īx* id. [? cf. Tlingit *ˀix*’ ‘to call; invite’ (Leer 1993)]
§77 Bur =*mál*- / =*mal*- ‘to be ashamed’ ~ Kott *amalaŋ* ‘fright’.

Again, it is possible that further study might enlarge this list. On the other hand, Bouda’s Burushaski-Caucasian comparisons fare much better, and of them the following have been incorporated into the current stock of DC etymologies:

§93 Bur *much* [*muś*] ‘snot’ ~ Chechen *march* [*mars*] id. (PEC **mHărčwV*).
§95 Bur -*umus* [=*ú-mus*, Y =*yú-mus*] ‘tongue’ ~ PNC **madz*- [Udi *muz*, etc. < **mĕlc̠ĭ*].
§103 Bur *tap* ‘leaf’ ~ Circassian (Adyge) *tħāp* [PNC **ƛ̣ăpi*].
§105 Bur *ti* [=*thi*-] ‘to pour’ ~ Avar *t’* [*t’e*-] id. [PEC **=V̄ṭwV*] [same etymology as Bouda (1964: §4), see below].
§114 Bur *taren*-um [*tharén*-um] ‘narrow’ ~ Avar *t’eren* [*ṭeréna*-b] ‘thin’ [PNC **=iƛ̣ɨ̆lV*; cf. Berger (1959, p. 26), below].
§118 Bur *ca* ~ Lak *ac’a* [=a=*ça*-] ‘to stand’ [PEC **HĕrçV*-]; but Archi -*occ’* [o=*c:i*-] < PEC **=Vmc̠Vr.*
§121 Bur *caghur* [*cháɣur*] ‘box for grain’ ~ Avar *caghur* [*caɣúr*] ‘corn bin, barn’ [PEC **cVɢVrV̆*].
§127 Bur *su-sun* [=*súsun*, Y =*sésen*] ‘elbow’ ~ Udi *sun* id. [PEC **s̠ɨ̄nŏ*].
§131 Bur *sui* [**=sú[m]*] ‘umbilical cord’ ~ Lak *c’u-n* [*çun*] ‘navel’, etc. [PEC **ʒŏnʔŭ*].
§135 Bur *tchi* ‘mir geben’ [**=ćhi*- ‘to give’ (class IV sing. obj.)] ~ EC **tcha* [**=ićV*] ‘to sell, distribute’ (Chamalal *ič*- ‘to sell, give’).
§139 Bur *butch* [*buć*] ‘ungelt male goat, 2 or 3 years old’ ~ Lezgi *bac’i* [*baçi*] ‘kid’, etc. [PEC **b[a]çV*].
§141 Bur *jiki* [**ʒ̭eké* (with retroflex initial)] ‘nit’ ~ Dargwa *c’ikwa* [**çikʷa* / **çekʷa*] ‘flea’, etc. [PNC **ć̣ăkwə̆*; cf. Bsq **siga-ṙ* ‘mite’].
§147 Bur *yachi* [*yaśi*-], *gicha* [*g=iśa*-] ‘to weave’ ~ PNC **chcha* [i.e. **šša* = NCED **=irś̠E* ‘to weave’: Avar =*es:*-, Tsezi =*iš*-, Dargwa Chirag =*erš*- / *ruš*-, etc.; cf. Bsq *josi* < **e=ośi* ‘to sew’].
§149 Bur *nach* [*naś*] ‘blame, reproach’ ~ Lak *natch* [*nač*], Avar *netch* [*nič*], etc. ‘shame’ [PEC **n[a]čV* or **n[a]ćV*; cf. Bsq **naś*-ka ‘repugnance, disgust, abomination’].
§153 Bur *chi* [*śi*] ‘hearth’ ~ PNC **c’a* [**çăjɨ̆*] ‘fire’ ~ Bsq *su* [**śu*] ‘fire’.
§164 Bur -*gha-n* [Y =*ɣán*, H N =*ɣə̆́n*] ‘heel’ ~ Avar *eghe* [*eʁé*] id. [PEC **ʔĭnGwV̆* < PSC **ʔĭGwVnV́*; cf. Bsq **hoin* ‘foot’].
§188 Bur *gatu* [*gaṭú*] ‘clothes; dress, garment’ ~ Avar *gorde* [*gordé*] ‘shirt’, etc. [PEC **gwĭrdwV*; cultural word – may have been borrowed independently].
§189 Bur *hun* ‘wood, timber’ ~ Chechen *hˡun* [*ħu*n] ‘forest’, Khwarshi *hun* ‘mountain’, Tsezi *xon* [*ħon*] id. [PNC **fănV*; cf. Bsq **oi(-)han* ‘forest, desert’].
§197 Bur *il* [Y =*híl*-, H N =*íl*-] ‘to dip, soak, wet’ ~ Chechen =*il* ‘to wash’ [PEC **=ăʎVr* ‘to wet, be wet, soak’].
§199 Bur *uy*-ám [Y *nyam*] ‘sweet, tasty’ ~ Avar *huin* [*huʔína*-b] ‘sweet, tasty’ [PEC **HnēHw*V̄; cf. Bsq **eme* ‘smooth, soft, slow’].

Bouda continued with articles in *Orbis* "Burushaski Etymologien" (1954) and "Burushaski Etymologien II" (1964). From the first we can cite:

> §1 Bur *bal* [causative =s=*pal*] 'to kindle' ~ WC *bl* [**bəlA*] 'burn, shine'.
> §4 Bur *darc* 'threshing floor' ~ Lak **darac'*- [*t:araça*-lu] id. [< PEC **ħrānʒū*].
> §14 Bur *ṣ̌au* [*ṣáo*] 'oversalted' ~ Circassian *š'əu* 'to salt', *š'əγ*^{o} 'salt' [Adyghe *ś̌əʁw*, Kabardian *šəʁw* < PWC **ʒ́ə* /**č́ə*; cf. Bsq **i=ća-śo* 'sea'].
> §19 Bur Y, N a-*ṣ̌ṭ*-iŋ [=*ṣṭ*-iŋ], H a-*ṣ̌ç̌*-iŋ [=*ṣç*-iŋ] 'loins, waist, kidney area' ~ Circassian *s*^{o}*tə* [**s*^{w}*t:ə́*] 'hind part, genitals ~ Basque **i=sta*- 'groin, 'thigh'].
> §22 Bur *gus* 'woman' ~ Dargwa *g*^{o}*aza* [*k:*^{w}*ac:a*], Lak *kk*^{o}*acca* [*k:*^{w}*ac:a*], Archi *g*^{o}*acci* [*g*^{w}*ac:i*]'mare' [< PNC **gwāʒ́ē*; Lak and Archi words are borrowed from Dargwa].[4]
> §30 Bur *γi*-n [H *γiín*, N *γáin*, Y *γēn*] 'thief', *γi*-ki [*γaí*-ki] 'theft' ~ Avar *=iq'i* [*=iq̇:*-] 'to steal', Dargwa *iɔ̤* [*=irʕ*^{w}- 'to deceive'] [< PNC **HirɢwĔ* - same etymology as Bouda (1950: §56), discussed above].

In "Burushaski Etymologien II" (1964), rather strangely, only *one* of the 39 North Caucasian etymologies offered seems to pass muster and is included in the current DC collection of etymologies, and it is essentially the same as one of his earlier comparisons: §4 Bur Y *hut*-in, H *ut*-in 'to fill, pour' [root is really *=thi*-] ~ PNC **t*^{o} [**=V̄ṭwV*] 'to pour, soak' [= the same comparison as Bouda (1950: §105), see above].

In Bouda's comparisons with Tibetan we now have to reject many, but note:

> §67 Bur **sin* [in Y *sésen*-um, *sesén*-um, H,N *sisín*-um] 'clean, clear' ~ [West] Tibetan *siŋs* 'clear, thin', b=*siŋ* 'thin, limpid (of fluids)' [cf. *saŋs* 'to clean, cleanse', (*b*)*seŋ* 'clean, white', etc. < PST **chĕŋ* 'clean, clear'; Bsq **susen* 'right, correct, just', **sin* 'oath, truth; true'; Chechen, Ingush *c'ena* 'clean, pure', Batsbi *c'ayn*, etc. < PNC **HăʒĔm*-].

Among the Bur-Tibetan comparisons we also find §69 Bur *šat* [*śat*] 'power, strength, force' ~ Tibetan *šed* [*śed*] id. It seems quite clear that the Bur word is Indo-Aryan, cf. Shina *śat*, Pali, Prakrit *satti,* etc. < OI *śaktí*- 'power' (Berger 1998.III: 392; CDIAL 12250). Since this word is apparently found only in Tibetan and Lolo-Burmese **šatx* 'strength, force', according to Peiros & Starostin (1996), it too may be borrowed from Indic. This comparison is the same as Bouda (1950: §17). Some other Bur-ST parallels may indicate areal borrowings in either direction:

- Bouda (1950: §12) Bur *duŋ* 'cunnus, vulva' ~ Tib. *ã̄=dom*-s 'genitalia'.
- Bouda (1950: §18) Bur *ta* 'leopard' ~ Tib. s=*tag* 'tiger' [Lepcha să=*thaŋ*].
- Bouda (1964: §66) Bur *chir*, Y *cer* 'line, row, turn, chain' ~ Tib. *chir* 'order, course, succession' [PST **chĭr*].

---

[4] Valid only if there was a semantic shift 'mare' > 'woman' in Bur. The Yen cognate **kuʔs* means 'horse' ('cow' in Ket).

Around the same time the Burushaski scholar Hermann Berger (1956, 1959) offered some interesting lexical comparisons between Burushaski and Basque.[5] As usual, some must be rejected, but yet some remain:

- Berger (1956, p. 7) Bur *hər* [*har*] 'corn worm' ~ Bsq *har* 'worm' [~ *ãr, haar, aar, ar* < **ha[m](a)r̄*; cf. PEC **ɦabarV* 'worm'].
- Berger (1956, p. 7) Bur *bər* [*bar*] 'valley' ~ Bsq i=*bar* 'valley'.
- Berger (1956, pp. 7, 10) Bur *galgí* 'wing' ~ Bsq *hegal* 'wing' [also 'loin, flank (of cow); fin (of fish)'; cf. PEC **q̱ɨ̆lʔi* 'elbow, arm, wing'].
- Berger (1956, p. 9, note 16) Bur *=l-pur* 'eyelid' [< **=l-t-* 'eye' + **bur* 'hair'] ~ Bsq *bepuru* 'eyebrow' [< **be(gi)*-t- 'eye' + **buru* 'head'].
- Berger (1956, p. 10) Bur *=ġan* [*=ɣán*] 'heel' ~ Bsq *oin, huin* [**hoin*] 'foot' [cf. Bouda (1950: §164), above].
- Berger (1956, p. 11, note 23) Bur *həranč̣* [*haránc̣*] 'wooden five-tined hayfork' ~ Bsq *arhe* 'harrow' [cf. PEC **ʁarhV* 'harrow'].
- Berger (1956, p. 12) Bur *-yon* [**yṓn*] 'all' ~ Bsq *honi* ['complete, complete number'; **hain*-ic 'many, much'].
- Berger (1956, p. 12) Bur *ġuraṣ̌* [*ɣuráṣ, ɣuriáṣ*] 'excrement' ~ Bsq *khorotz* [**Koroc*] 'dung' [cf. PEC **ḳ[u]rč̣V* 'dung, excrement'].
- Berger (1956, p. 16) Bur *hólalas*, Y *ahúlal* 'moth, butterfly' [**haúlal*-] ~ Bsq *euli, uli* [**e=uli*] 'fly'.
- Berger (1956, p. 17) Bur *ṣ̌oq* [*ṣoq*] 'sole of shoe' ~ Bsq **ośki* 'shoe' [cf. PEC **šälḵwV* 'hoof'; or **šwŏq̱̇HwV* ~ **q̱̇HwŏšwV* 'heel, ankle'?].
- Berger (1956, p. 18) Bur *-yeš* [*=yeéś*] 'permanent residence' ~ Bsq *etŝe* [**e=će*] 'house, cottage' [cf. PNC **c̣[ī]ju* 'house'].
- Berger (1956, p. 22) Bur *nok* 'bent, crooked (human limb)' ~ Bsq *nokhu* [**noku*] 'physical defect' [if the latter is not < Romance: cf. Lat. *nocuus* 'hurtful, noxious'].
- Berger (1956, p. 24) Bur *čhaṣ̌* [*ćhaṣ*] 'thorn bush' ['sand thorn, brier'] ~ Bsq *śaśi* id. ['bramble, thorn, bramble bush, thicket'; cf. PNC **ʒāʒĕ* 'thorn, prick'; PY **seʔs* 'larch'].
- Berger (1959, p. 26) Bur *bién*-um 'narrow, thin, slender' ~ Bsq *mehe* [< **behe*-n?] 'thin' (inanimate) [cf. PST **pā* 'thin'].
- Berger (1959, p. 26) Bur *tarén*-um [*tharén*-um] 'narrow' ~ Bsq *lirain* 'slender, svelte, lithe' [cf. PNC **=iƛ̣̆ɨlV* 'thin' = Bouda (1950: §114), above].
- Berger (1959, p. 26, note 34) Bur *tar*-íŋ 'skin bag or bottle' ~ Bsq *larru* [**laṙu*] 'skin, leather' [PNC **Ł̠ŏli* 'colour, to paint' ('skin, sheepskin' in Dargwa)].
- Berger (1959, p. 26, note 34) Bur *ter* 'mountain pasture, summer pasture' ~ Bsq *larre* [**laṙe*] 'pasture, meadow' [cf. PEC **ŁwĕłV* 'enclosure, fence'; PST **răl* 'fence, framework'].[6]
- Berger (1959, p. 27, note 35) Bur *gišáyas, yašias* [root *g=*iś*- / *=y=*áś*-] 'to weave, knit' ~ Bsq *josi* [**e=ośi* 'to sew'] [cf. PNC **=irś̠E* 'to weave' = Bouda (1950: §147), above].
- Berger (1959, p. 27, note 35) Bur Y *holtaṣ̌, hultaṣ̌* [*hultáṣ*] 'barefoot' ~ Bsq *orthuts, ortotz* [**ortúć*] 'barefoot' [related to PEC **=ōmŁ̠V* 'to put on (trousers, shoes)'].

---

[5] In a note (Berger 1959, p. 17-18, note 3) he joined Burushaski and Basque in a family with – *Dravidian*(!).

[6] This cognate chain exemplifies a common change: 'fence' > 'enclosed space' > 'meadow', 'field', etc.

- Berger (1959, p. 33, note 57) Bur *-múltur* ‘nostril’ ~ Bsq *muthur* [**muturṙ < *murturṙ*] ‘snout, muzzle; end, edge’ [cf. PEC **wĕnƛ̣V* ‘beak, horn, head’ > Batsbi *marƛ́ŏ* ‘nose’, etc.].
- Berger (1959, p. 34, note 57) Bur *tur, =ltur* ‘horn’ ~ Bsq *adar* [**adaṙ < *a=rdaṙ*] ‘horn’ [cf. PEC **ƛ̱wĭrV* ‘horn, braid, mane’ > Avar *ƛ:ar* ‘horn’, etc.].
- Berger (1959, p. 34, note 57) Bur *šuɣurí* ‘a kind of hard pear’ ~ Bsq *sagar* [**śagar̄*] ‘apple’ [cf. Lezgi *č:üχʷer* ‘pear’, etc.].[7]
- Berger (1959, p. 36) Bur *halánc* ‘moon’ ~ Bsq *ilazki* [< **hila-s-ki* ?] ‘moon‘.
- Berger (1959, p. 36) Bur *gunc* ‘day’ ~ Bsq *iguzki* [< **egun-s-ki* or **Higun-s-ki* ? ‘sun’: cf. Bsq **egu-n* ‘day’, PY **xiʔG* ‘day’, **xiGa* ‘sun’, PEC **Hwīq̇V* ‘day’, etc.].
- Berger (1959, p. 37) Bur *har* ‘valley, dale’, pl. *harm*-iŋ ~ Bsq *haran [*ɦaran*] ‘valley’ [cf. PEC **ʔārV* ‘plain’; PY **ʔ[ā]r*$_1$*V* or **h[ā]r*$_1$*V* ‘taiga, wilderness’].
- Berger (1959, p. 37) Bur *hunám* ‘fine cloth (as wedding gift), bridal ornament’ ~ Bsq *ehun* ‘cloth’ [**e=hun*, cf. **eihu-n* ‘to weave, braid’, PEC **=irχ̣wVn* ‘to knit, weave, spin’].
- Berger (1959, p. 40, note 79) Bur **hur* ‘water’ (now ‘water conduit, trough’) [also H *hurtá* ‘wet’, *hurú*-ginas ‘wave, stream, whirlpool’, N *hurú*-ṣ ‘becoming wet, overripe’, H, N *hurú* ‘juice of overripe fruits’] ~ Bsq **hur* ‘water’ [cf. PY **xur*$_1$ ‘water’; PEC **ħwirɨ* ‘lake, pond’; alternatively Bsq **hur* can be compared with PEC **ɦwɨ̆lV* ~ **lɨ̆ɦwV* ‘river, reservoir’].

Besides the lexical parallels, Berger noticed some regular and non-trivial correspondences, namely Bur initial *t-* = Bsq initial *l-* (see the three comparisons beginning with Bur *tarén*-um “narrow”), and Bur intervocalic *-lt-* = Bsq intervocalic *-rt-/-rd-* (the comparisons “barefoot”, “nostril”, and “horn”). In current DC phonology these are reckoned as reflexes of original lateral affricates **Ł* / **ƛ* /**ƛ̣* (see **3. Phonology**). After producing what can only be called strong contributions to the evidence for the DC hypothesis, later in life Berger abandoned the connection, or at least considered it unprovable (Berger 1992: 6).[8]

Olivier Guy Tailleur (1958, 1994) included Burushaski and Yeniseian in his “Paleo-Eurasian,” a superstock that also included Basque, Caucasian (incl. Kartvelian), and even Ainu! Nevertheless, he remarked upon some important DC morphological parallels:[9]

---

[7] Lezgi *č:üχʷer* ‘pear’, Tabasaran *žiχir*, Agul *č:iχˤar*, Tsakhur *ǯɨχˤā*, Kryz *ǯüħür*, Budukh *čöħür*; beside Rutul *χˤɨr*, Archi *χˤert*, Udi *ar* ‘pear’; cf. Chechen, Ingush *qor* ‘pear’, Batsbi *qor* ‘apple’; other cognates in Lak and Dargwa, all < PEC **qHŭ̄re* or **qŭ̄Rʔe.* NCED tries to explain the words with initial sibilant affricates [*ǯ, ž, č:, č*]: “As for the strange first part **ǯu-* … , it is most probable that we deal with a compound **ǯum-χIera* "quince"+"pear" with the first component distorted.” Alternatively, a word such as *čVχʷVr* ‘pear’ or ‘apple’ (supported by external comparison with Bur *šuɣurí* and Bsq **śagar̄*) could have existed alongside **qHŭ̄re* or **qŭ̄Rʔe,* was lost outside of Lezgian, and influence of **ǯum* ‘quince’ was secondary.

[8] “...konnte eine Verwandtschaft ihrer Sprache mit irgendeiner anderen bisher nicht nachgewiesen werden; die strukturelle Ähnlichkeit mit dem Baskischen und den Kaukasussprachen ist auffallend, aber es fehlt an überzeugenden Wortgleichungen selbst im Bereich der Körperteile, Verwandtschaftsnamen und niederen Zahlwörter.”

[9] Some forms have been updated.

- (1994: p. 39) Ket and Yug *-d-* in oblique cases ~ Basque **-t-* (as in *su-**t**-argi* 'firelight' ['fire' + **-t-* + 'light']) ~ various Caucasian and Kartvelian case morphs. [Tailleur interpreted this **d* as a genitive morpheme; we now see it as an oblique stem marker. Cf. Burushaski **-t-* in some plural forms such as (H, N) *guśpúr-**t**-aro* 'princes', (Y) *=yúhar-**t**-iŋ* 'husbands'; and compounds such as *=l-**t**-éṣ* 'eyebrow' (cf. Lak *i-**t:a**-c'ani* id.]
- (1994: p. 40) Ket *-s, -as, -aś* [instrumental, comitative], Kott *-oš / -aš* [comitative] ~ Basque *-z* [instrumental] ~ Chechen *-sa*, Ingush, Batsbi *-s* [ergative animate], Dargwa *-s* [dative], etc. < PNC *-s:- [instrumental animate; general attributive]. [Cf. Burushaski *-as / -áas* [verbal infinitive], e. g., *hér-as* 'to cry, wail', *min-áas* 'to drink, smoke'. The development to [infinitive] is convergent with that in some Lezgian languages, e.g. Lezgi *-z* [dative], [infinitive].]
- (1994: p. 41) Yug *-ey* (in locative *-k-ey / -g-ey*); fossilized in locutions such as Yug *xɔt-éy* '(to the) front', *fič-éy* 'down(stairs)', Northern Ket *hɨt-e*[6] 'down(stairs)', Ket *akś-ey*[5] 'why?', Kott *peg-ai* 'elsewhere', etc. ~ Basque *-i* [dative] ~ Avar *-e* [dative], *-n-e* [infinitive], etc. < PEC **-Hi* [dative ?]. [Cf. Burushaski *-e* [ergative and genitive],[10] e. g. (H, N) *hiles-e* 'boy' (erg.), *hiles-e* 'boy's', etc.]
- (1994: p. 41) Ket *-ka / -ga* [locative inanimate], Kott *-iga* [dative inanimate] ~ Basque *-ik* [partitive], *-k* [ergative] ~ Northwest Caucasian **-k'V* [instrumental, comitative, etc.]. [Cf. Burushaski [instrumental] *-k /-ak*: *thúr-ak* "with a whip", *hundó-k – bundó-k* "with sticks and stones", etc. (Berger 1998: I, 61).][11]
- (1994: pp. 41-2) Ket *-ŋa* [dative] (in I-class *-da-ŋa* / II-III *-di-ŋa*, etc.) = Yug *-ŋ* (I *-da-ŋ* / II *-di-ŋ*, etc.) ~ Basque *-n* [locative], as in *etxe-an* 'at the house, in the house', *kale-an* 'on the street', etc.; also *-en* [genitive], as in *gizon-ar-en* 'of the man' ~ Chechen *-n*, Lezgi *-n*, Udi *–n* [genitive] < PNC **-nV* [genitive], also shifted to dative, ablative in other languages. (Note that PSC **ŋ* has merged with **n* in Basque and Caucasian.) [Cf. Burushaski (Y) *-(i)ŋa* [comitative] 'together with', e.g. *yék-iŋa* 'with the name ..' ; (H) fossilized *-aŋe* [instrumental], (N) *-aŋe, -iŋe, -eŋe* id.]
- (1994: pp. 42-3) Kott *-čaŋ* [ablative, elative], e.g. *ūl-i-čaŋ* 'out of the water' ~ Basque**-ca-t*, e.g. *seme-tzat daukat* 'I consider him my son', *zoro-tzat naukan* 'je le tenais pour fou'; (possibly also**-c* in other agglutinated case forms: **-(r)a-n-c* [directional], e.g. *mendirantz* 'toward the mountain'; Northern Basque **-(r)a-ko-c* [destinative]). [Cf. Burushaski (Y) *-ći* [general locative], e.g. *=rén-ći* 'in the hand'; (H, N) fossilized [locative], e.g. *hála-ći* 'in the goal (in polo)'.]
- (1994: pp. 43-6) Gender (class) systems that distinguish Yeniseian, Burushaski, and (some) Caucasian languages from surrounding Eurasian languages, and some material parallels in class marking among the three families discussed. The Ket vowels *-u-* (~ *-a-, -e-*) I-class (masculine) vs. *-i-* (~ *-ɨ-, -ə-*) II-class (feminine) correspond to some degree with the Northeast Caucasian **u̯=* and **y=*, respectively, most clearly in the object person markers *-a-* ~ *-o-* I-class / *-i(d)-* II-class. (Strangely, in Burushaski these are reversed as *=i-* I-class / *=u-* II-class.) Tailleur also calls attention to the Ket inanimate (III-class) object marker *-b-*, corresponding to Northeast Caucasian **w=* ~ **b=* (inanimate III-class).

---

[10] Except class II (feminine), where the genitive ending is *-mu* (Berger 1998: I, 58).

[11] This suffix is lexicalized and only occurs with certain nouns and with verbs meaning "strike" or "shoot".

- (1994: p. 46) Tailleur relates the Ket III-class [inanimate] predicate adjective suffix *-am*, e.g. *śuĺem-am* 'red', to the Burushaski lexicalized adjective suffix *-um / -úm*, as in *bur-úm* 'white', *ćhaɣúr-um* 'cold'. In addition he cites the Lak adjectival suffix *-ma, -mä, -mi, -mur*, as in *luhe-mä* 'that which is black' (NCED *luħi-s:a* 'black').
- (1994: p. 46) Plural morphemes. Yeniseian plurals *-n* and *-ŋ* are compared with *n*-plurals in Ubykh, Batsbi, Dargwa, Khinalug, Hurrian, Urartian, and Burushaski *-ŋ* in noun plurals and *-n* in verbal plurals.
- (1994: p. 49) Verbal "ruiniforms": this term, borrowed by Tailleur from geology, aptly describes the tendency in DC (particularly in the 'core' group, i.e. Caucasian, Burushaski, and Yeniseian) for verbal stems to wear down and truncate to very short forms, e.g. Burushaski **=t-* 'to do, make', Yeniseian **di* 'to lie down, put down', PNC **=ătVr* 'to let, leave; to stay'. Since this is also true of Kartvelian (e.g. PK **gw-* 'to feed') and other languages with extensive prefixation and suffixation, it is more a typological or areal feature than strictly genetic.
- (1994: pp. 50-1) Personal markers: notes some of the pronominal parallels already discussed by other scholars, e.g. Ket *ad* 'I' ~ Avar *du*-n, *di*-, etc.; Ket *ū* 'thou' ~ Abkhaz *w*=ara, Archi *u*-n, Burushaski *u*-n, etc.; Ket *ku*= 'thou' ~ Basque *h*=, Burushaski *gu*= / *go*=, etc. These and more are discussed in detail in the **Morphology** section below.
- (1994: pp. 51-5) Remarks on various verbal morphemes, such as present marker *-*y*- in Yeniseian and Burushaski (the latter with verb stems in *-n*, e.g. *=min-́* 'to drink' : pres. *=miy-́*); past tense markers *-*n*- and *-r_1- in Yeniseian ~ Basque *-n* (*neki-**en*** 'I knew it'); Burushaski ***nu**=ku=ći-**n*** 'having given it to thee' [absolutive or converb]; PNC **-nV* [past gerund / participle], e.g. Hunzib *uhu-**n*** 'having died', Chechen *w=aχ-**ana*** 'he has left', Lezgi *t'ü-**na*** 'I have eaten', etc.

More recently, George van Driem (2001) has endorsed the Burusho-Yeniseian link under the name 'Karasuk', mainly based on grammatical homologies.

Beginning in the 1970-80s, some members of the "Moscow School" picked up the threads left by earlier scholars. Vladimir N. Toporov, a prolific and wide-ranging Russian philologist, wrote an article (Toporov 1971) that laid some of the theoretical groundwork for the Sino-Caucasian hypothesis of the next decade.

> The existence of genealogical connections between the Yeniseian languages and those of the Tibeto-Burmese linguistic family … are regarded now already as doubtless … [and] if the given convergences are taken for reliable, the Basque, North Caucasian and Burushaski languages might be regarded as islets, having remained from a formerly indivisible chain of languages stretched in the latitude direction, from the Atlantic deep into Central Asia … the Yeniseian languages … might be regarded as the next remaining link of the chain …

After paying homage to earlier efforts by several scholars (Trombetti, Bouda, Holmer, Tailleur, Dulson) Toporov then went on to focus on the connections between Burushaski and Yeniseian. After mentioning some typological similarities (phonetic and morphological, such as the noun class systems and polysynthetic verbal morphology), he outlined a significant number of material parallels, some of which can be summarized in the form of a table (with updated reconstructions):

Table 1: Burusho-Yeniseian Grammatical Comparisons by Toporov (1971)

		Burushaski	Yeniseian
Noun: plural suffixes	Inanimate Plural ***-ŋ**	IV-class **-ŋ**[12]	Ket III-class **-ŋ**[13]
Noun: possessive prefixes		(Y) **a**=rén, (H, N) **a**=ríiŋ 'my hand' (Y) **gu**=rén, (H, N)**gu**=ríiŋ 'thy hand' **gú**=mi 'thy mother'	Ket **ab**=ĺaŋ 'my hand', **ug**=ĺaŋ 'thy hand' Yug **k**=am 'thy mother'
Pronoun stems	1sg ***(ʔa)ŋ(a)**[14]	**a**=	***b**= / ***ʔab**=
	2sg ***(ʔu)K(u)**	**gu**= / **go**=	***kV**= / ***ʔV)k**=
	3sg ***mV**	**mu**= (II=class)	***wV** > Ket **bu**
	Deictic ***ki-**	(Y) **khin, khené** 'this' (I-class) / **khomó** (II) / **gusé, khos** (III) / **guté, khot** (IV)	Ket **kīdə** 'this' (I-class) / **kida**^{6} (II, III)[15]
Verbal subject	Plural ***-n**	**-n**	***-n**

Toporov's younger Moscow colleague Sergei Starostin, who had worked on the reconstruction of Proto-North Caucasian, Proto-Yeniseian, and Old Chinese, re-examined many of these earlier proposals and concluded that the genetic affinity of Caucasian, Yeniseian, and Sino-Tibetan, at least, could be demonstrated on the basis of shared basic vocabulary and grammar, confirmed by regular phonological correspondences. His first articles describing what he called "Sino-Caucasian" were published in 1982 and 1984. The 1982 article outlined Starostin's reconstruction of Proto-Yeniseian and proposed etymologies connecting Yeniseian, Caucasian, and Sino-Tibetan, many of which reiterated the earlier comparisons of Trombetti, Bouda, and others, but this time the classic Indo-Europeanist method was applied, proposing regular sound correspondences among the three proto-languages. The 1984 article (1991 in English) expanded on this and included some comparisons not involving Yeniseian as well. It will not be necessary to repeat any examples here, since (with a few exceptions and revisions) they are essentially identical with the most current etymologies and correspondences found in Starostin (2005a, 2005b) and the Tower of Babel databases.

Another Muscovite, Ilia Peiros (1988), offered lexical comparisons between Burushaski and Starostin's Sino-Caucasian. Besides repeating Bleichsteiner's pronoun comparisons, and Bouda's comparisons of 'eat', 'leaf', 'umbilical', 'goat', 'clothes',

---

[12] With various vowel and consonant onsets.

[13] Exceptions are usually phonetically conditioned: nouns ending in *–n* take the ending *–ŋ* regardless of class, and nouns ending in *–ŋ* take the ending *–n* (Werner 1994).

[14] These are not Toporov's reconstructions, but those formulated later by Starostin et al.

[15] Pakulikha and Surgutikha dialects.

'oversalted', discussed above, Peiros discovered some new parallels, some of which have been adopted in the current hypothesis:[16]

- §3. Bur *hol* 'army, troops' // PY **har-* 'slave, servant; Arin' // PNC **ʔwə̆hri* 'troops, army' // PST **rāH* 'enemy, captive, war' [cf. Basque **heṙi* 'town, inhabited place, people, nation'; a common semantic shift of 'warrior' > 'captured warrior' > 'slave'].[17]
- §5. Bur *=l-ći(-n)* 'eye' // PY **de-* // PNC **ʔwĭlʔi*.
- §10. Bur *qoś* 'dwelling house, quarters, lodging' // PY **χuʔs* 'tent (made of birch bark)' > 'house' // PEC **q'wIΛ[rc]Λ* 'house'. [It turns out that "PEC **q'wIΛ[rc]Λ*' does not exist, as explained by Starostin (1995a: 305); cf. instead PEC **q̇VmčV* 'balcony, verandah' ?]
- §14. Bur (H, N) *asií* 'star' // PNC **ʒwhărī / *ʒwăhrī* 'star' // PST **sēŋ* 'star' [As it turns out, the PNC word does not belong here, but goes with Bsq **i=saṙ* instead. The underlying Bur form seems to be **a=sVm*, based on (Y) *asúmun*, and the (H, N) plurals *asií-m-uc*, and also matches Na-Dene: PA **səŋʷʔ* = **səmʔ* 'star'.]
- §30. Bur *ćak* 'pickaxe' // PY **čok* 'axe' // PEC **č̱wɨḳwV* 'knife' // PST **ʒ́hV̄kʷ* 'axe, chisel, hoe'.
- §32. Bur *=san* 'spleen' // PY **seŋ* 'liver' // PNC **c̱wăjmĕ* 'gall, anger' // PST **sĭn* 'liver' [A classic DC comparison, now extended to Basque *beha-*su[m]* 'bile, gall';[18] and Na-Dene: PAE **=sən-t'* 'liver' (Vajda 2010)].
- §38. Bur *sis ~ ses* 'persons, people' // PY **ǯeʔ*-ŋ 'people' // PNC **ćwĭjo* 'man, male' (Ubykh *ćə́ća* 'persons, people').
- §45. Bur *jawá* (**ʒ́awá*) 'thick sinew' // PEC **ć̱wiwV* 'muscle'.
- §50. Bur *=ltúr* 'horn' // PEC **ƛ̱wɨ̆rV* 'horn, mane, braid' [thus expanding Berger's comparison, discussed above, of Bur *=ltúr* and Bsq **a=daṙ* < **a=rdaṙ* 'horn'].
- §63. Bur **ʒ́ām* 'distant relative' // PEC **ʒ́ămV / *maʒ́V* 'relative, kinsman'.
- §65. Bur **=ltul-* 'to saddle', **tilí-* 'saddle' // PEC **ƛ̱wiłē* 'saddle' [the comparison is semantically and phonetically perfect, though one wonders if horse riding is early enough to exist at the time of Proto-Western- Dene-Sino-Caucasian (ca. 6000 BCE). See the discussion below].

A paper by Blažek & Bengtson, "Lexica Dene-Caucasica" (1995), was the first to assemble a large number (219) of Dene-Caucasian lexical and grammatical etymologies that included all six families included in the current hypothesis: Basque, North Caucasian, Burushaski, Yeniseian, Sino-Tibetan, and Na-Dene.

For more recent work on Burushaski and Dene-Caucasian see the **Postscript**, below. In the rest of this paper the lexical and grammatical evidence assembled over a span of more than eight decades is re-examined in the light of the most recent developments.

---

[16] Transcriptions and reconstructions have been updated/normalized.

[17] With a seemingly unique (irregular) correspondence of Bur **l* = PSC **r*. This is not surprising considering the wavering between *r ~ l* in Old Indic (an areal neighbor of Bur), as in *lih- ~ rih-* 'to lick', *labh- ~ rabh-* 'to seize', *róman ~ lóman* 'hair', etc.

[18] This word looks like an old compound, where the first element **beha-* ('liver'?) could correspond to PST **phe* 'spleen' (e.g., Thankur *=pəy* 'liver'); PY **b[a]jbVl* 'kidney' could be a similar compound < **b[a]j-pɨʔɨĺ* (where the second element seems to be an altered form of PY **pɨʔɨĺ* 'intestines').

Morphological evidence: Pronouns

Some of the Dene-Caucasian morphological evidence discussed above, and more, can be summarized as follows:

Pronouns: Both Burushaski and the reconstructed Proto-(North) Caucasian have *suppletive* pronoun stems in the first and second person singular; and, in our analysis, the original suppletive stems themselves are inherited, though with some paradigmatic rearrangements. According to Nikolayev & Starostin (NCED pp. 402, 483-84, 855, 1014-15, 1084-85), the original Proto-Caucasian paradigms were very complicated, and difficult to reconstruct with much certainty. For the present purpose, let us compare Hunza Burushaski (Berger 1998, vol. I, p. 80) with two East Caucasian languages, Khinalug and Tsakhur (and PEC):

Table 2: Burushaski and East Caucasian Pronouns

		direct	**ergative**	**genitive**	**dative / oblique**
1st person singular ‘I’ / ‘me’	Burushaski	ʒ́e[19]	ʒ́á-a	ʒ́áa	áa-r [dat.]
	Khinalug	zɨ	yä	i ~ e	as [dat.]
	Tsakhur	zɨ	za	yiz-in	za- [obl.]
	PEC	*zō(-n) ~ *nĭ	*ʔez(V)	*ʔiz(V)	*zā- [obl.]
2nd person singular ‘thou’ / ‘thee’	Burushaski	un ~ uŋ ~ um	un-e	góo	góo-r [dat.]
	Khinalug	wɨ	wa	wi	oχ [dat.]
	Tsakhur	wu ~ ʁu[20]		yɨʁ-na, yɨʁɨ-n[21]	wa- [obl.]
	PEC	*u̯ō(-n) ~ *ʁwV̄	*ʔŏʁwV	*ʔiu̯V-	*ʔŏʁwV [obl.] ~ *dū- [dat.]

Clearly a great deal of rearrangement has taken place in these languages since the postulated original paradigms of thousands of years ago. Nevertheless it is possible to posit the following suppletive pronoun stems:

- First person singular (1): PDC ***ʐV*** ‘I / me’ > Bur *ʒ́e*, *ʒ́á-*; Khinalug, Tsakhur *zɨ*, etc. (PNC **zō*).
- First person singular (2): PDC ***ŋV*** ‘I / me’ > Bur *áa*-r [dat.] ‘to me’, also as pronominal prefixes *a=́* / *á=* / *áa=* ‘I, my’ (i.e. *a* < **ŋa*, a regular change in Bur: see SCP 48-49). This stem does not figure in the Khinalug and Tsakhur forms above, but only in Lak (*na* ‘I’)

---

[19] In Indological tradition conventionally written *je, jáa*, etc.

[20] According to NCED Tsakhur *wu* ‘thou’ (< **uō*) and *ʁu* ‘thou’ (<**ʁwV̄*) are in free variation.

[21] *yɨʁ-na* for 1- and 2-class, *yɨʁɨ-n* for 3-class.

and Dargi (*nu* ‘I’, *nu*-ša ‘we’, etc.). PDC **ŋ* regularly becomes **n* in PNC/PEC **nĭ* ‘I’ and Basque **ni* ‘I’.

- Second person singular (1): PDC ***wV*** ‘thou / thee’ > Bur *u*-n; Khinalug *wɨ*, Tsakhur *wu*, etc. (PNC **u̯ō*). Some Cauc forms come from the suffixed PEC **u̯ō-n*, similar to Bur *u-n*, such as Archi, Udi *un* ‘thou’, Lezgi, Agul *wun*, Avar *mun*, Andi *min* ‘thou’ (assimilated < **bin* < **win*), etc.
- Second person singular (2): PDC ***xɢwV*** ‘thou / thee’ > Bur *góo, góo*-r, also as a pronominal prefix, with many shapes depending on the following stem (*gu=́ / gú= / -kú= / gó= / -kó= / góo= / -kóo=*); Khinalug *oχ* [dat.], Tsakhur *ʁu* [direct], *yɨʁ*-na, *yɨʁɨ*-n [gen.]; elsewhere in Cauc: Chechen, Ingush, Batsbi *ħo* ‘thou’, Dargi *ħu* ‘thou’ (dial. *gu, ʕṵ, ṵ, i*),[22] Rutul (dial.) *ʁu*, Udi (dial.) *hu*-n ‘thou’; and Basque **hi* ‘thou’ (restricted intimate use), verbal affixes **h*=, *-*k* < *-*ga*.

In Yeniseian analogous suppletions are evident:

- PDC ***zV*** ‘I / me’$_1$ > PY ***ʔaʒ*** ‘I’ > Ket *āt*, Kott *aj* (*ay*), Pumpokol *ad*, etc.;
- PDC ***ŋV*** ‘I / me’$_2$ > PY ***b-/*ʔab-/*-ŋ*** > Ket *āp* ‘my’, *ba=/bo=* [1st pers. sg. verbal affix], Kott *-ŋ* [1st pers. verbal affix] (for phonetics see SCP 48-49);
- PDC ***wV*** ‘thou / thee’$_1$ PY ***ʔaw / *ʔu*** > Ket *ū* ‘thou’, Yug *u,* Kott, Arin *au*, Pumpokol *úe* id.;
- PDC ***xɢwV*** ‘thou / thee’$_2$ > PY ***kV- / *ʔVk-*** > Ket *uk* ~ *ūk* ‘thy’, *k=, ku=* [2nd pers. sg. verbal affix], etc.

In other Dene-Caucasian subgroups the original suppletive paradigms have been regularized or leveled in various ways. In West Caucasian only the morphs **zō* ‘I / me’ and **u̯ō* ‘thou / thee’ have survived (PWC **sa* and **wa*, respectively). In Basque, on the other hand, PDC **zV* and **wV* have been eliminated, and only the stems **ŋV* and **xɢwV* remain, as **ni* ‘I’ and **hi* ‘thou’,[23] respectively. Purely by chance the East Caucasian language Dargi has rearranged the first and second person direct forms to coincide with those of Basque: Dargi (Akusha, Urakhi) *nu* ‘I’ / *ħu* ‘thou’ = Bsq **ni* / **hi*.

An interesting discussion of Dene-Caucasian pronouns, including Sino-Tibetan and Na-Dene forms as well, has been provided by George Starostin (2010a).

Lexical evidence:
Burushaski-Caucasian-Basque-Yeniseian comparisons[24]

Body part terms

---

[22] /ṵ/ represents a pharyngealized vowel, caused by the Proto-Dargi “emphatic laryngeal” **ħ*.

[23] The other, less intimate and more frequent Basque word for ‘thou’, *zu* (**su*), is a recent development of the original 2nd person plural pronoun (= PNC **źwĕ* > Chechen *šu*, Lak *zu*, Archi *žʷe*-n, etc. ‘you’ [pl.]), analogous to the development in English in which *thou* has been mostly supplanted by the original plural *you*.

[24] In each etymological entry, the Burushaski word or words are listed first, with dialectal designations (H = Hunza, N = Nager, Y = Yasin) where appropriate. Next, proposed Caucasian cognates, if any, then Basque and Yeniseian cognates, if any, and finally there may be notes, following the symbol §. The notes may include more remote Sino-Tibetan and Na-Dene cognates.

Bur. **bác-in* > *bácin* (H, N) ‘shank; (animal’s) hind leg above the hock’[25]
~ Yeniseian: PY **baʔt-* ‘knee’ > Ket *bat*-puĺ^5, Arin karam-*pat* ‘elbow’, *pat*-as ‘knee’, etc.
~ Cauc: Avar *púrc:i* ‘ham’, Chamali *bec*^{w} ‘thigh; knee (of animal)’, Tabasaran *bac* ‘paw’, Tsezi *besi* ‘fist’, etc. < PEC **b[ə]c̱V* (NCED 291)
~ Basque **borc-/*bośt-* ‘five’ < ‘*hand’ (cf. Tsezi ‘fist’, above).
§ Cf. PST **pŭt(-s)* ‘knee’ < PDC **bV́stV* (SCG 19-20).

Bur. **búmbal-*, in *búmbalten* (Y) ‘ankle’ (old compound with **=lten* ‘bone’)[26]
~ Yeniseian: PY **bul* ‘foot, leg’ > Ket *būĺ*, Kott *pul* ‘foot, leg’, *bul*-aŋ ‘on foot’, Arin *pil* ‘feet, legs’, etc.
~ Cauc: PEC **bĭmłV̆* ‘hoof, foot’ (> Tsez *bula* ‘hoof’, Chechen *ber-g* id., Avar *mal* ‘foot’, etc.: NCED 307) + PEC **ƛ̱wVnʔV* (see the entry for Bur. **ltén* ‘bone’, below).
§ Cf. PST **phə̆l* ‘upper part of foot (leg)’ < PDC **bĭmłV̆́* ‘foot’ (SCG 14-15).

Bur. **bur* > *bur* (H, N, Y) ‘(single) hair’, =l-*pur* (H), =r-*pur* (N), =mú-r-*puli*-aŋ[27] (Y) ‘eyelid’ (where =*l-* ~ =*r-* = ‘eye’), =*pur*-aŋ (H), =ṣ-*pur*-aŋ (N), =ṣ-*puri*-aŋ (Y) ‘mane (of animals)’ (where =*ṣ-* = ‘neck’)[28]
~ Basque **buru* ‘head’, *bet-*puru* ‘eyebrow’ (**be(t)-* = **begi* ‘eye’).
§ Bur. and Basque compared by Berger (1956, p. 9, note 16). PDC **burV* ‘hair’ (Starostin 2005d, 2007 [TOB]).[29]

Bur. **galgi* > *galgí* (H, N, Y) ‘wing, fin’
~ Cauc: Lak *qa* ‘wing’, Lezgi *ʁil* hand’, Bezhta *χaro* ‘elbow’, etc. < PEC **q̱ĭlʔi* ‘wing, arm, elbow’ (NCED 895)
~ Basque **ma=gal* ‘wing’, **e=gal* ‘wing, fin’ (northern Bsq *hegal* apparently influenced by **hega-* ‘to fly’).
§ Bur. and Basque compared by Berger (1956: 7). PDC **xqwV̆́lʔi* (SCG 235).

Bur. **γā́n* > =*γáan* (H, N), =*γán* (Y) ‘heel’[30]
~ Cauc: Avar *eʁé* ‘heel’, Lak *niq:a* ‘heel’, Chechen *ħaq*-olg ‘ankle, ankle-bone’, etc. < PEC **ʔĭnɢwV̆* ‘heel; ankle, ankle-bone’ (NCED 248)
~ Basque **hoin* ‘foot’.

---

[25] **c* or /c/ in these and all Dene-Caucasian words henceforth, is consistently the sibilant affricate [t^s].
[26] /=/ at the beginning of a morpheme indicates a bound morpheme that requires a possessive prefix (in Burushaski) or a class prefix (in Caucasian). Postulated fossilized prefixes (as in Basque) are also followed by this symbol, e.g. Bur. (H) *a=ltín* ‘my bone’, *mu=ltín* ‘her bone’; Avar *w=as* ‘son’, *y=as* ‘daughter’, *b=as* ‘young (of animal)’; Basque **bi=hoc* ‘heart’, etc.
[27] *-aŋ* is a common plural ending.
[28] /ṣ/ denotes a retroflex sibilant.
[29] The PDC form would have to be **buŕV*, with a palatalized or ‘soft’ rhotic (SCP 60), otherwise the Bsq word would be **buŕu* (**burru*), with a trilled rhotic.
[30] Burushaski /ɣ/, written *γ* and later *ġ* by Berger, is really a voiced uvular fricative /ʁ/ and is part of the back-velar or uvular series with /q/ and /qh/ (the latter often realized as [q^χ] or just [χ]), though in realization the velars and uvulars overlap to some extent (Berger 1998 I, pp. 20-21).

§ Bur + Avar compared by Bouda (1950, §164). Bur + Basque compared by Berger (1956: 10). All < PDC **ʔĭɢwVnV́* 'heel, foot' (SCG 265).[31] Bsq development could have been **ʔĭnɢwV̆ > *ʔwĭŋ(V) > *hoin.*

Bur. **huṭ > =húṭes* (Y), *=úṭis, =úṭ* (H, N) 'foot'[32]

~ Cauc: Avar *ħet'é / ħet'* 'foot', Chechen *t'a* 'front leg (of animal)', Dargi Urakhi, Sirgokala *t'ʷaħ* 'foot', etc. < PEC **ṭwīħV̄ ~ *ħīṭwV̄* 'foot, forefoot' (NCED 1007).[33]

§ Bleichsteiner (1930, §32) compared Bur + Avar. Cf. PST **tĭH* 'heel, ankle' < PDC **ħīṭwV̆ ~ *ṭwīħV̄* (SCG 207), or possibly **ɦwīṭV̆* would better account for the Bur form.

Bur. **ken > =́kin* (H, N), *=́ken* (Y) 'liver'[34]

~ Cauc: Andi *k':unu* 'kidney', Chamali *k'ũm, k'ũ̄*, Bezhta *koma* id., etc. < PEC **ḵunHV* (NCED 728).

§ Cf. PST **gjV̆nH* 'kidney' < PDC **(x)ḳúnHV* 'kidney' (Starostin 2005d, 2007 [TOB]). Na-Dene: Eyak *q'əma·* 'kidney, salmon roe'; PA **q'u·nʔ* 'roe'.[35]

Bur. **kur- > kúur* (H) 'finger-joint, toe-joint'; *kuróŋ* (H) 'bone'

~ Cauc: Chechen *k'uram* 'bone (for playing dice)', Lezgi *k'ur* 'hoof, leg (of animal)', Archi *k'ʷiri* 'leg (of animal)', etc. < PEC **ḳwĭrV* (NCED 736).

§ Cf. PST **k(h)rey* 'foot' < PDC **ḳwĭrV́* 'leg' (SCG 123); Na-Dene: PA **qe·-* 'foot', Eyak *=qi-/=qe·-* id. (Sino-Dene 20; there is no **r* in older levels of Na-Dene).

Bur. **ltén > ten* (Y) 'bone', *tan-c* 'leg'; *=ltín, tin* (H, N) 'bone', *=ltán-c* 'leg'

~ Cauc: Avar *tł':an* 'groin', Archi *tł':on*-t'ol 'fingernail',[36] Agul *k:un* 'ankle', etc. < PEC **ƛ̱wVnʔV* (NCED 785).[37]

§ Cf. PST **ləŋ* 'shin, ankle' < PDC **ƛ̣wV́nʔV* 'ankle, shin' (SCG 139-140).

Bur. **ltur > tur* (Y), *=ltúr, tur* (H, N) 'horn'

~ Cauc: Avar *tł:ar*, Chechen *kur* 'horn', Lezgi *firi* 'mane', etc. < PEC **ƛ̱wɨ̆rV* 'horn; braid, mane' (NCED 771)

~ Basque **a=daṙ* 'horn' (< **a=rdaṙ*).

§ Bur + Bsq compared by Berger (1959, p. 34, note 57); Bur + PEC compared by Peiros (1988, §50). PDC **ƛwɨ̆́rV* 'horn' (SCG 134-135).

---

[31] **ɢ* or /ɢ/ in PDC, PNC, PEC, and PY reconstructions represents a voiced uvular stop, though possibly realized as an affricate [ɢʁ], and reflexes in Cauc and Bur are often the simple fricative /ʁ/ ~ /ɣ/.

[32] Confusingly, **ṭ* or /ṭ/ denotes a retroflex stop in Bur, but a glottalized stop /t'/ in PDC, PNC, PEC. Similarly all underdotted consonants in Bur are retroflex, but glottalics in PDC, PNC, PEC.

[33] Metathetic variants are quite frequent in Cauc languages.

[34] The notation /=́/ indicates that the obligatory pronominal prefix, rather than the noun stem, bears the accent, e.g. (H) *á=kin* 'my liver', *gó=kin* 'thy liver', etc.

[35] For typology of 'kidney ~ liver' cf. Old Indic *vr̥kká* 'kidneys' > German Romani *pukko* 'liver, lungs, spleen, kidney', Oriya *buku* 'heart, chest, courage', etc. (CDIAL 12064).

[36] Archi *-t'ol* < PEC **ṭĭłʔV* 'finger'.

[37] The PNC/PEC sound **ƛ̣* is a glottalized lateral affricate, alternatively written /tł'/, as in Navajo.

Bur. *_melc̣_ > =_mélc̣_ (H, N, Y) ‘jaw, jawbone’[38]

~ Yeniseian: PY *_binč-_ > Ket _bint_ ‘face’, Kott _púnǯ_-ol ‘chin’, Arin _pinǯ_-al ‘chin’, etc.[39]

~ Cauc: Archi _muč_ ‘nose, beak; peak, top’, Lak _murč:i_ ‘lip, brim (of dish)’, Chechen _mʕäč-ig_ ‘tip (of something) bent upwards’ < PNC *_mħwɨ̆lc̱i_ ‘edge, tip’ (NCED 816).

§ PDC *_mħw[ɨ́]lči_ ‘face, cheek, jaw’ (SCG 145). Cf. Na-Dene: PA *=_wə̂š-_ / =*_wə̂č’_- ‘cheek’, Eyak =_waš_ id.

Bur. *_meṣ_ > =́_miṣ_ (H, N), =́_meṣ_ (Y) ‘finger, toe’

~ Yeniseian: PY *_bes-_ > Ket _bɛs_-taq^5 ‘index finger’ (compound with PY *_təʔq_ ‘finger’).

~ Cauc: Kryz _mič’_-ek ‘hoof, nail, claw’, Lak x:i-_mič’_ ‘hoof’, etc. < PEC *_mič̣V_ ~ *_mic̣V_ (NCED 819).

§ PDC *_Hmeć̣V́_ (SCG 77).

Bur. *_moq-_ > =_móq-iṣ_ (H, N) ‘cheek’, =_móq-iṣ_ (Y) ‘face’, =_móq-ot_ ‘cheek’

~ Dargi _biq’_-ri ‘witness’, Avar _nuʕ_ ‘witness’ (<*_nuq̇V_ < *_muq̇V_), Chechen _baq’_ ‘true’ (adv.), Ingush _boq’o_ ‘truth, rule’, etc. < PEC *_wĭmq̇V_ ‘witness, true’ (< ‘*eye’) (NCED 1050)[40]

~ Basque *_moko_ ‘beak; extremity, point; face; façade’, etc.

§ Cf. PST *_myV̆k_ ‘eye’ < PDC *_wĕmq̇V́_ (SCG 216); Na-Dene: Tlingit _wàɢ_ = _wàq_ ‘eye’, PA *_-nə-**wēɢ**-əʔ_ ‘eye’; Eyak _lāχ_ ‘eye’ < *_nāχ_. Proto-Na-Dene, like Proto-Yeniseian, probably had no initial *_m-_, which seems to have changed to *_w-_ (see Sino-Dene, p. 225). For Bur & Bsq phonetic development (*_wĕm-_ > *_mo-_) cf. the next set (*_wĕn-_ > Bur & Bsq *_mu-_).

Bur. *_múltur_ > =_múltur_ (H, N) ‘nostril’

~ Cauc: Bezhta, Hunzib _motł’o_ ‘beak’, Hinukh _mitł’u,_ Khwarshi _matł’u_ id., etc., Batsbi _marło_ ‘nose’, Karata _butł’a_ ‘horn’, etc. < PEC *_wĕnƛ̣V_ (NCED 1041)

~ Basque *_mutuṙ_ ‘snout, muzzle’ < *_murtu-ṙ._

§ Bur. and Basque compared by Berger (1959, p. 33, note 57). PDC *_wĕnƛ̣V́_ (SCG 216). For Bur & Bsq phonetic development (*_wĕn-_ > *_mu-_) cf. the foregoing set.

Bur. *_muś_ ‘nose; end, edge’ > =_múś_ (Y) ‘nose; end, edge’, =_múś-puṭ_ ‘beak’; _muś_ ‘end, edge’ (H, N), =_múpuṣ_ ‘nose’

~ Cauc: Chechen, Ingush _muts’ar_ ‘snout, muzzle, trunk’, Avar _móts’:u_ ‘teat, nipple; tip’, Lezgi _murz_ ‘blade; edge, verge; narrow side of an object’, etc. < PEC *_mħə̆rc̱ū_ ‘protruding part, point, edge’ (NCED 811)

~ Basque *_mośu_ > Gip _musu_ ‘nose’; in other dialects: ‘snout, face, lip, kiss,’ etc.

§ A different comparison, with PEC *_mħăc̱ĕ_ ‘edge’ (Ingush _mʕiz_-arg ‘snout’, etc.) in Starostin (2007 = TOB).

---

[38] /c̣/ represents a retroflex affricate.

[39] “PY, however, has no initial *m-: in this position we observe *w- [in pronominal stems only], *p- [the general reflex] or *b- [from initial clusters *Hm-, *mH-]” (SCP 30-31).

[40] For semantic typology of ‘witness ~ eye’, cf. Old Indic _sākṣin-_ ‘witness’ : _sākṣāt_ ‘with the eye, clearly’ < _sa-_ ‘with’ + _akṣa-_ = _akṣi-_ ‘eye’ (Buck 21.23). That ‘eye’ is the archaic meaning is shown by the Eastern Dene-Caucasian (= “Sino-Dene”) words, cited above.

Bur. *_p(h)at-_ > _=́pat_ (H), _=́phat_ (N), _=́p(h)at_ (Y) ‘side, flank’
~ Cauc: Lezgi _p:ad_ ‘side’, Kryz _badow_ ‘near’, Budukh _bode_ ‘near’ < PEC *_bVdV_ ‘side’; cf. Urartian _bedə_ ‘side; (postpos.) on the part of, by’ (NCED 315)
~ Basque *_pata-ṙ_ ‘slope, rugged slope’ < ‘*side’.
§ A root with two stops, showing probable assimilations and dissimilations: *_patV_ ~ *_padV_ ~ *_badV_. Reconstructed as PDC *_ṗVdV_ (SCG 165-166).

Bur. *_phol-_ > _phol-ɣó_ (Y), _phul-ɣúuỵ_ (H), _phur-ɣúuỵ_ (N) ‘feather’ (compound with *_ɣuỵ_ ‘hair’)
~ Cauc: Lak _p’iħulli_ ‘feather’, Dargi (Akushi) _pạħạla,_ Chechen _pēla-g_ id., etc. < PNC *_ṗVħVɫV_ (NCED 879)
~ Basque *_bilho_ ‘hair, mane’ (not derivable from Latin _pilum_).[41]
§ PDC *_ṗVħVɫV_ (SCG 166), or metathesized *_ṗVɫ(V)ħV._

Bur. *_=́qat_ > _=́qat_ (H), _=́qhat_ (N), _=́qet_-araŋ (Y) ‘armpit’
~ Yeniseian: PY *_qot-_ (~ *_χot-_) > Ket _qɔta_^{6} ‘in front, before’, Yug _xɔtéy_ ‘(to the) front’, Arin ún-_kut_ ‘in front of’, etc.; adverbial development of the noun: cf. Eng. _abreast_, etc.
~ Cauc: Avar me-_héd_ ‘brisket (chest of animal)’, Bezhta _ʁade_ ‘brisket’ < PEC *_qVdV_ (NCED 897).
§ PDC *_qVdV́_ (~ *_xqVdV́_) ‘breast’ (SCG 170). Cf. Na-Dene: Navajo _=ɣìd_ ‘chest’, Tlingit _χèt_ ‘chest, breast’ (Boas), _xe`t-ka_ ‘beastbone’ (Leer 1993); Haida _sq’ut_ ‘armpit’.

Bur. *_=qhaś-_ > _=qháś_-iŋ (H, N) ‘hind end, arse’, _=xáś_-aŋ (Y) ‘female sex organ’[42]
~ Cauc: Udi _qoš_ ‘behind’, a development of PNC *_-Vq̱V_ ‘behind’ (NCED 1026).
§ Comparison by Bouda (1954, p. 229, no. 28).

Bur. *_=qhát_ > _=qhát_ (H, N), _=xát, =xat_ (Y) ‘mouth’
~ Cauc: Lak _qịt’_ ~ _q_^{w}_ịt’_ ‘Adam’s apple, beak’, Udi _χirtäy_ ‘Adam’s apple’, Kryz _χulut’_ ‘larynx’ (< *_χut’_-ul), etc. < PEC *_qwīṭi_ (NCED 905).
§ PDC *_qwə̄ṭí_ ‘palate, mouth’ (SCG 172).

Bur. *_=qhorpVt_ > _=qhúrpat_ (H, N), _=xórpet_ (Y) ‘lung’
~ Cauc: Andi _χuns:ir_ ‘lung’, Tsezi _χot’ori,_ Archi _χurt:ur_-t:i, Dargi Chiragh _qusara_, Akusha _χurhala_, _χurhari_ id., etc. < PEC *_q̱wələV(rV̄) / *χ̱wələV(rV̄)_ (NCED 901)[43]
~ Basque *_hauśpo_ > (Bzk) _auspo, aspo_ ‘lungs; bellows’, only ‘bellows’ in other dialects. The Bsq word seems to be contaminated with *_hauć_ ‘ashes, dust’ (as if ‘dust blower’).
§ Bur. and Basque compared by Berger (1959: 21). The words in the three families have a certain resemblance to each other, but seem impossible to reduce to a single protoform.

---

[41] Latin _pilu_ would become Bsq *_biru_ or *_piru_, according to the development of other Lat loanwords in Bsq, e.g. Bsq _goru_ ‘distaff < Lat _colu(m)_, etc. Further, the laryngeal cluster /lh/ in Bsq is explainable by PDC *_ṗVɫ(V)ħV._ The IE status of Lat _pilum_ is itself questionable.

[42] _-iŋ, -aŋ_ are frequent plural affixes, seen in several of these etymologies.

[43] “The root is rather peculiar (although there are no doubts at all in its EC antiquity): it contains a very rare phoneme *-ɵ- (which yields t-like reflexes in PTs, Lak. and PL, but s-like reflexes in And. and Darg.)” (NCED).

Bur. =*sán* > =*sán* (H, N, Y) 'spleen'

~ Yeniseian: PY **seŋ* 'liver' > Ket *śēŋ* 'liver', Yug *seŋ* id., Arin *šen*-óugbu 'heart', etc.

~ Cauc: Avar *c:in* 'gall, anger', Tindi *s:imi* id., Archi *s:am* 'gall', Lezgi *seb*, Bezhta *simo* id., etc. < PNC **c̱wăymĕ* 'gall; anger' (NCED 329)

~ Basque *beHa-*sun* 'bile' (apparently an old compound).[44]

§ Peiros (1988, §32) compared Bur + PNC + PY + PST **sĭn* 'liver'. PDC **cwắyŋĕ* 'liver, bile' (SCG 22-23). Cf. also Na-Dene: PAE **sənt'* 'liver' (Vajda 2010).

Bur. *=*sú[m]* > =*sú* (Y: plural =*súmu*), =*súi* (H, N: plural =*súimuc*) 'navel, umbilical cord'

~ Cauc: Dargwa Chirag *zu* 'navel', Khinalug *c'um*, Tindi *c:ũ̄*, Lak *c'un*, Chechen *c'on*-ga id., etc. < PEC **ʒŏnʔŭ* (NCED 1096).

§ Bouda (1950, §131) compared Bur + Lak, etc. PDC **ʒŏ́nʔŭ* (SCG 249). Cf. Na-Dene: Eyak *c'a·ʔ* 'umbilical cord' (< **c'ãʔ*); PA **c'e·q'* 'navel' (cf. Chechen *c'on-ga* < **c'an-k'u*).

Bur. *=*sV́sVn* > =*súsun* (H, N), =*sésen* (Y) 'elbow'

~ Cauc: Udi *sun* 'elbow', Rutul *sɨn* 'front part of leg', Lak *s:an* 'foreleg, paw, pad', Avar *san* 'organ, body part', etc. < PEC **s̱īnŏ* (NCED 963)

~ Basque **san*-ko 'leg, calf, foot', **san*-ka-ŕ 'tibia, heel-bone', etc.

§ Bouda (1948, §100) compared Bsq + Lak; Bouda (1950, §127) compared Bur + Udi. PDC **sə̄́nŏ* (SCG 187-188).

Bur. **tal* > *tal* (H) 'stomach, belly'

~ Cauc: Avar *t'ul* 'liver', Andi *relitł:i,* Lak *t:ilik'* 'liver', etc. < PEC **Hƛ̣älV* ~ **Hläƛ̣V* 'liver' (NCED 586).

§ Note recurrent correspondences of Bur initial **t-* with Caucasian lateral affricates (**ƛ*, **ƛ̣*, **Ł*) and Basque **l-*: see the next two entries, Bur **ltén* 'bone' and **ltur* 'horn' (above), and the **Phonology** section of this paper. PDC **Hƛ̣älV́* 'liver, belly' (SCG 76).

Bur. **táno* > *táno* (H, N) 'colon, rectum'; probably related to *táno, tanéelo* 'illegitimate child, of low birth'

~ Cauc: Avar *t'ínu* 'bottom', Tindi *hinlu*, Bezhta *õtł'o*, Lezgi *k'an* id., Khinalug *k'an*-ɨk' 'under', etc. < PNC **Hƛ̣ŏnŭ* (NCED 590).

Bur. **tar*[*í*]- '*skin' > *tar-íŋ* (H, N, Y) ' bag made from animal hide (for containing fluids, or for rafts)'

~ Cauc: Avar *tł':er* 'color' (< *'skin'), Karata *tł':ere* id., Dargi Akusha *guli* 'skin, hide, sheepskin', Tabasaran *ɣal* 'bark, shell', etc. < PNC **Ḻŏli* '*skin' (NCED 789)

~ Basque **laṙu* 'skin, hide, leather'.

§ Bur and Bsq compared by Berger (1959, p. 26, note 34); Bsq and Avar compared by Bouda (1948, §86). PDC **Łŭ́li* 'skin, color' (SCG 130; I find the PST and PY comparanda here very doubtful). But based on the evidence of outliers (Bsq & Bur) for internal **-r-* the Cauc forms may have assimilated the laterals (**Ƚ̱ŏri* > **Ƚ̱ŏli*), and there may have been vowel metathesis, ie. **Ƚ̱ŏru* > Bsq **laṙu*, **Ƚ̱ŏri* > Bur **tar*[*í-*]. Cf. a similar situation with Bur **ter* ‘high pasture’ ~ PEC **ȽwĕłV* ‘fence, yard’ ~ Bsq **laṙe* ‘pasture, grassland’ (below).

---

[44] For **beHa-*, cf. more distant DC relatives: PST **phe* 'spleen' (Thankur =*pəy* 'liver', etc.); PY **b[a]jbVl* 'kidney' could be a similar compound < **b[a]j-pɨʔɨĺ* (where the second element = **pɨʔɨĺ* 'intestines').

Bur. **=úl* > *=úl* (H, N, Y) ‘belly, abdomen, bowels’

~ Cauc: Tindi b=*etł’:u* ‘stomach, rennet, abomasum’ (*b=* is a class prefix), Dargi Akusha *=arg* ‘stomach, inside’ (with changing class prefixes), Agul *uray* ‘rennet, abomasum’, Rutil *ɣɨriɣ* id., etc. < PEC **=īraȽ̱V* (NCED 670)
~ Basque **urdail* ‘stomach’ (Bzk also ‘rennet; womb’).
§ PDC **=ī́rȽV* ‘stomach, belly’ (SCG 112-112).

Bur. **=yáldi-* > *=yáldir* (H), *=yáldin* (N) ‘the part of the ribs under the armpit; middle part of the breast’

~ Cauc: Tindi *hatł’:ar* ‘armful’, Chamali *hatł’:*^{w}, Hunzib *hetłu* id., etc. < PEC **Hĕƛ̱V ~ *HĕȽ̱V* (NCED 558).
§ For the regular correspondence of Burushaski *-lt- (-ld-)* to Caucasian **-ƛ̣-*, **-Ƚ-*, see the **Phonology** section of this paper. PDC **HV́ƛ̣V* ‘armful, armpit’ (Starostin 2005d, 2007 [TOB]).

Bur. **ʒ́al- / *ʒ̣al-* > *jaláalimiŋ* (H, N) ‘long hair (of people)’, *jaláṣ* (Y) ‘hairy’, *j̣aléi, j̣alíi* (H) ‘beard (of goat)’, etc.[45]

~ Cauc: Karata *žale* ‘mane’, Lak *zulū* ‘nap, pile’, Bezhta *žaro* ‘horse’s mane’, etc. < PEC **ʒ́ăłhī* (NCED 1101).
§ PDC **ʒ́ăłhī́* ‘mane, long hair’ (SCG 251).

Nature

Bur. **baŋ* > *baŋ* (N), *baŋgí* (Y) ‘resin (of trees), pitch, gum’

~ Cauc: Chechen *baga* ‘pine tree’, Lezgi *muk’*-rag ‘fir tree’, Lak (dial.) *mịk’ịk’ịy* ‘pine cone’, etc. < PEC **bħĭnḵwV* ‘pine tree’ (NCED 296)
~ Basque **muki* > (Bzt) *muki* ‘gum, resin of trees’ (with regular change of **bVnk-* > **mVk-*).
§ PDC **bħĕ́nḳwV* ‘pine tree, resin’ (SCG 13). Cf. Old Indic *bhaṅgá* ‘hemp, bhang, marijuana’ (a resinous plant) > Hindi *bhā̃g*, *bhaṅg*, etc., possibly one of the Burushic words acquired during the Indo-Aryan sojourn in the Northwest (CDIAL 9354; Witzel 1999, p. 4).

Bur. **bar* > *bar* (H, N, Y) ‘small valley, ravine, gorge’

---

[45] *ʒ̣, /j̣/ represent a voiced retroflex affricate, alternatively written /dẓ̌/.

~ Basque **i=baṙ* ‘valley’.

§ Bur. and Basque compared by Berger (1956: 7; 1959: p. 28, note 39). Apparently an old DC word that was lost in Caucasian but preserved in the outliers.

Bur. **bun[d]-* > *bun,* (pl.) *bundó* (H, N) ‘mountain pasture, mountain grove; boulder; wild, mountain-’, *bun,* (pl.) *bundó ~ bunjó* (Y) ‘boulder’

~ Cauc: Khinalug *mɨda ~ mda* ‘mountain’ (cf. Georgian *mta* ‘mountain’)

~ Basque **mendi* ‘mountain’.

§ Bur. and Basque compared by Berger (1959, p. 28, note 41). Possibly another archaic DC word (cf. preceding set) that was lost in Cauc (except in the outlier Khinalug). Georgian *mta* is also isolated within Kartvelian and may be borrowed from archaic Cauc.

Bur. **dul-dúm* > *duldúm* (H, N) ‘rising cloud (of dust, smoke, etc.)’

~ Cauc: Archi *diƚ*^{w} ‘cloud’, Lak *t:urlu* ‘cloud’, Dargi Akusha *dirix* ‘cloud’, etc. < PEC **dilʎ̱wV* (NCED 400).

§ PDC **dílʎwV* ‘dust, cloud’ (SCG 40).

Bur. **ge* > *ge ~ gye* (H, N, Y) ‘snow’

~ Yeniseian: PY **qo* (~ **χo*) ‘ice’ > Ket *qō* ‘ice’, Yug *xō* id.

~ Cauc: Khinalug *q:i* ‘cold’ (n.), Lak -ạ-*q:i*- ‘to grow cold, catch cold’, Andi =eʁ-eš- ‘to get cold, freeze’, etc. < PNC **=HĭɢĀ* ‘to freeze, get cold, be benumbed’ (NCED 568).

§ PDC **=HĭɢĀ́* ‘ice, to freeze’ (Starostin 2005d, 2007 [TOB]).

Bur. **γónder-* > *γónderes ~ γondoles* (Y) ‘water, that runs over many stones’

~ Cauc: Botlikh *ʁadaru* ‘stream, brook’, Godoberi *ʁada* ‘ravine’, Lak *ạtara* ‘mountain stream’, Chechen *ʕowr-aš* ‘mountain stream (after rain or thaw)’, etc. < PEC **ʁHwadVrV* (NCED 478).

§ PDC **ʁHwadV́* (SCG 185). Bur *-n-* is unclear.

Bur. **hun* > *hun* (H, N) ‘wood, timber, beam, hewn trunk’, *hun* (Y) ‘wood, firewood’

~ Cauc: Chechen *ħun* ‘forest’, Chadakolob *xuná*-q ‘shady side, slope’, Khwarshi *hun* ‘mountain’, etc. < PNC **f̄ănV* ‘mountain, hill’ (NCED 425)

~ Basque: **oihan* (*oi=*han*) ‘forest, woods; mountain; desert’.

§ Bouda (1950, §189) compared Bur + Cauc. PDC **xŋwănV́* ‘height, mountain’ (SCG 232). This etymology exemplifies the common shift of ‘mountain, hill’ > ‘forest, woods’: cf. Old Slavic *gora* ‘mountain’ ~ Lithuanian *girià* ‘forest’, etc. (Buck 1.22, 1.41).

Bur. **ltap* > *tap* (H, N, Y) ‘leaf, petal; leaf (page) of book’; *=ltápu-* (H, N), *=ltápi-* (Y) ‘to wither’

~ Yeniseian: PY **yə̄pe* ‘leaf’ > Ket ʌ:, Yug ʌ:*hp*, Kott *d́ipi*, etc. ‘leaf’

~ Cauc: Chamali (dial.) *lapã* ‘leaf’, Lak *č’ap’i,* Dargi Akusha *k’api*, Adyge *tħāp* ‘leaf’, etc. < PNC **ƛ̣ăpi* ‘leaf’ (NCED 774)

~ Basque **lapa-ṙ* ‘bramble, thorn’.

§ Bouda (1950, §103) compared Bur + Cauc. PDC **ƛ̣ăpí* ‘leaf’ (SCG 136). With the common correspondence of Bur **t-, *=lt-, *-lt-* ~ PNC **ƛ̣*: see **Phonology**.

Bur. *_mal_ > _mal_ (H, N, Y) 'field'

~ Cauc: Archi _mał:i_ 'winter pasture', Avar _márxi_ 'farmstead', Lak _maš:i_ 'farmstead', etc. < PNC *_malx̱wē_ (NCED 795).

§ PDC *_malx̱wē_ 'pasture' (Starostin 2005d, 2007 [TOB]). Indo-Aryan origin of the Bur word (Sanskrit _māla_- 'forest near a village', Prakrit _māla_- 'garden', etc. CDIAL 10088) does not seem likely.

Bur. *_phét-_ > _phét-iŋ_ (H, N, Y) 'ashes'

~ Cauc: Chamali _bat'a_ 'faeces', Tabasaran _bat'ur_ 'dirt', Ingush (dial.) _bɛd_ 'faeces', etc. < PEC *_bHāṭV_ 'dirt, faeces' (NCED 299).

§ PDC *_ṗHāṭV_ 'dirt' (Starostin 2005d, 2007 [TOB]). For typology of 'ashes ~ dirt ~ excrement' cf. Old Indic _kṣayá_ 'loss, waste' > Panjabi, Hindi _kheh_ 'ashes, dust, rubbish, ordure', etc. (CDIAL 3661); OI _pāṁśu_ 'crumbling soil, dust, sand; dung, manure', etc. > Bengali _pā̃s_ 'ashes', Awadhi (dial.) _pā̃si_ 'manure', Hindi _pā̃s_ 'dust, dung', etc. (CDIAL 8019); and the set including Bur *_ther_ 'dirt', below.

Bur. *_phunc_ > _phunc_ (H, N, Y) 'dew'

~ Yeniseian: PY *_piʔt_ 'glue' > Ket _hiʔt_, Yug _fiʔt_ 'glue', Kott _fīt_ ōginaŋ 'to glue'

~ Cauc: Lak _pic'_ 'dew, sweat', Avar _pic':_ 'resin', Dargi _penc'_ 'resin', Ubykh _bzə_ 'water', etc. < PNC *_pĭnc̱wĂ_ 'resin, juice' (NCED 871)

~ Basque *_pista_ > (Bzk) _pizta_ 'fresh rheum (secretion from eyes)'.

§ PDC *_pĭ́nsṭwĂ_ 'glue, resin' (SCG 160-161). But surely the original PDC sense was 'secretion, issue of liquid', whether of human or animal body (> 'rheum, sweat'), of trees (> 'resin, gum, pitch' > 'glue'), or of nature ('dew' > 'water'), etc.

Bur. *_śi_ > _śi_ (H, N) 'fireplace, hearth'

~ Cauc: Ingush _c'i_ 'fire', Lak _c'u,_ Avar _c'a_, Godoberi _c'ayi_, Bezhta _c'o_, Abkhaz á=m-_ca_ 'fire', etc. < PNC *_c̣ăyɨ̆_ 'fire' (NCED 354)

~ Basque *_śu_ / *_i=ću_ 'fire' > (c) _su_ 'fire'; Araban _itsu arri_ 'flint' ('fire-stone').

§ Bur + Cauc + Bsq compared by Bouda (1950, §153); Bur + Bsq compared by Berger (1956: 17). PDC *_c̣ăyɨ̆_ (SCG 23).

Bur. *_śVre_ '*night' > _gó(i)n-śare_ (H, N), _gón-śere_ (Y) 'the whole night, all the night through' (_gon_ = 'dawn')

~ Cauc: Avar _sor_-dó 'night', Chechen _süyrē_ 'evening', Batsbi _psara_ 'evening', _psare_ 'yesterday', Ubykh _šʷəwá_ 'night', etc. < PNC *_śwĔrV_ 'evening' (NCED 977).

§ PDC *_śwĔrV́_ 'evening', but with a different Bur cognate in SCG (194).

Bur. *_taɣ_ > _taɣ_ (Y) 'branch, shoot'

~ Cauc: Hinukh _tł'iχ_ 'bough', Avar _t'eχ_ 'flour (made of young sprouts), oat flour', Lak _k'urχ_ 'sprout', etc. < PEC *_ƛ̣ōrχwV_ 'sprout' (NCED 780).

§ With the common correspondence of Bur *_t-_ ~ PNC *_ƛ̣-_: see next three entries, and **Phonology**.

Bur. *_tápi_ > _tápi_ (H, N) 'stone terrace'
~ Cauc: Chadakolob _t'eb_ 'millstone, whetstone', Avar _tł'eb_ 'stone', Chechen _laba_ 'shed; peak (of cap)', etc. < PEC *_ƛ̣ĕṗV̄_ 'stone plate, shed' (NCED 777)
~ Basque *_lape_ > (Zub) _lape_ 'shelter under the eaves of a shed'.
§ PDC _ƛ̣ĭṗV̄_ 'flat; slab' (SCG 137). With the common correspondence of Bur *_t-_ ~ PNC *_ƛ̣-_: see **Phonology**.

Bur. *_ter_ > _ter_ (H, N, Y) 'high pasture, summer mountain pasture'
~ Cauc: Archi _tłoli_ 'yard, place in front of the house', Avar _lol_ 'open enclosure (for sheep)', Bezhta _kalo_ 'fence', etc. < PEC *_ŁwĕłV_ 'enclosure, fence' (NCED 791)
~ Basque *_laṙe_ 'grassland, pasture'.
§ Bur. and Basque compared by Berger (1959, p. 26, note 34). PDC *_ŁwĕłV́_ 'fence' (SCG 130, not including the Bur word), but this is virtually the same as the PEC form which, in comparison with the outliers Bur and Bsq, exhibits lateral assimilation (PDC *_Ł[w]ĕrV́_ > PEC *_ŁwĕłV_). Cf. a similar situation with Bur *_tar_[_í_]- '*skin' ~ PEC *_Ł̱ŏli_ '*skin' ~ Bsq *_laṙu_ 'skin, leather' (above).

Bur. *_tiṣ_ > _tiṣ_ (H, N, Y) 'wind'
~ Cauc: Tsezi _łaci_ 'wind', Akhwakh _tłãc'o_ 'voice, shout', Tindi _łač:u_ id., etc. < PEC *_ƛ[a]rʒ́V_ 'movement of air' (NCED 767).
§ PDC *_ƛV́rʒ́V_ 'wind' (SCG 134). With the common correspondence of Bur *_t-_ ~ PNC * _ƛ-_: see **Phonology**.

Bur. *_tumá-ỵ_ > _tumáỵ_ (H, N) 'nutshell, shell of fruit pit', _tumá_ (Y) 'hard shell (of nut, egg), fruit stone'
~ Cauc: Chechen _t'um_ 'marrow; kernel (of a fruit, nut)', Archi _t'ummul_ 'grape', Abkhaz a-_t'amá_ 'peach', etc. < PNC *_ṭŭmhV_ (NCED 1004).
§ PDC *_ṭŭmhV_ 'kernel (of fruit), seed' (SCG 205). Bur. *_tumá-ỵ_ and Lezgian *_ṭum(:)-ul_ (> Archi _t'ummul_, etc.) exhibit similar suffixation. Bur *_ỵ_ is often of lateral origin (see Bengtson & Blažek 2011a).

Bur. *_ther_ > _ther_ (H, Y), _ther-k_ (N) 'dirt'
~ Cauc: Akhwakh _tere-ti_ 'ashes, dust', Tindi _tira_ 'dung (of dog)', Bezhta _tär_ 'dung (of sheep)', etc. < PEC *_türV_ (NCED 993)
~ Yeniseian *_dəʔər_$_1$- 'powder, dirt' > Kott _taran_ 'dirt', etc.
§ PDC *_tV́rV̆_ 'dirt, dust, powder' (SCG 201). For typology of 'ashes ~ dirt ~ excrement' cf. Bur *_phét-_, above.

Bur. *_yaltar_ > _yáltar_ (H, N) 'upper leafy branches, crown (of tree)'; cf. _galtár_ (H, N, Y) 'small branch, twig'
~ Cauc: Avar _ʕartł':él_ 'branch, bough', Tsezi _atł'iru_ 'pod', Bezhta _ãtł'alo_ id., etc. < PEC *_ħălƛ̱VłV_ 'branch, pod' (NCED 508)

~ Basque **adaṙ* 'branch, knot (of tree), leg (of chair, bed)' < **ardaṙ*: a homonym of **adaṙ* 'horn'.

§ For the regular correspondence of Burushaski **-lt-* to Caucasian **-ƛ̣-* see the **Phonology** section of this paper.

Wild animals

Bur. **ć(h)argé > ćargé* (Y) 'flying squirrel'

~ Yeniseian: PY **saʔqa* 'squirrel' > Ket *saʔq*, Yug *saʔq / saʔx*, Kott *šaga,* Arin *sava*, Pumpokol *tak* 'squirrel'

~ Cauc: Chechen *šat'q'a* 'weasel', Andi *sartł':u* 'weasel', Tsakhur *sok* 'weasel', Adyge *cəʁʷa* 'mouse', etc. < PNC **cārgwV* (NCED 322)

~ Basque **śagu* 'mouse' > (c) *sagu*; in compounds: *sagu-zahar* 'bat' ('mouse-old'); *sat-or* 'mole' (< **śag-t-hor* 'mouse-dog'?), *sat-itsu* 'shrew' (< **śag-t-iću* 'mouse-blind').[46]

§ PDC **[c]ā́rxgwV̆* 'squirrel, weasel' (SCG 21). The original initial consonant is uncertain, probably altered here or there by expressive changes (cf. Bsq *sagu* /śagu/ 'mouse' ~ *xagu* /šagu/ 'mousie, wee mouse'). The meaning 'squirrel', only in Bur and Yen, may be evidence for their common development in a "Burusho-Yeniseian" subgroup of DC. (Cf. Bur **khen* 'flea' ~ PY **qəʔń* 'flea', below.)

Bur. **c̣hen > c̣hin* (H, N), *c̣en* (Y) 'small) bird'

~ Cauc: Abaza *c'i-s* 'small bird, sparrow', Avar *ħinč':* 'bird', Dargi Akusha *hunuc'* 'eagle (poetic)', Chechen *ħōza* 'sparrow', Archi *noc'* '(small) bird, sparrow', etc. < PNC **ɦnīc̱(w)ĭ ~ *ɦc̱(w)īnĭ* 'small bird' (NCED 525)

~ Basque **hunc* 'owl'.

§ The comparison assumes metathesis: PDC **ɦnī́[ć̣]ĭ ~ *ɦ[ć̣]ī́nĭ* 'bird' (Starostin 2005d, 2007 [TOB]), apparently also **ɦ(w)ī́n[ć̣]ĭ* to account for Dargi *hunuc'* 'eagle' and Bsq **hunc* 'owl'.

Bur. **γórku- > γúrqun* (H), *γúrquc* (N), *γórkun* (Y) 'frog'

~ Cauc: Tindi *q':orq':u,* Khinalug *q'urq'or,* Khwarshi *q'urq'ač* 'lizard', Kabardian (ħandər-)*q:ʷāq:ʷa* 'frog', etc. < PNC **q̱̇wVrVq̇V̄* (NCED 942).

§ Bleichsteiner (1930, § 21) compared Bur + Cauc. Obviously of onomatopoeic origin, but Bur and Cauc forms exhibit precisely matching forms. PDC **xq̇wV́rV(q̇V̄)* 'frog' (SCG 243).

Bur. **har > har* (H, N) 'corn worm, grain weevil'

~ Cauc: Avar *ħapára* 'worm', Andi *habara*, Tsakhur *ạbra*-wuč'e id. < PEC **ɦabarV* (NCED 508)

~ Basque **ha*[*m*]*aṙ* 'worm' > Rnc *ãr*, Lap *har*, archaic Bzk *haar* 'worm', etc.

§ Bur. and Basque compared by Berger (1956: 7). PDC **ɦaṗárV ~ *ɦabárV* 'a kind of worm' (SCG 91). Some forms apparently altered by expressive changes, e.g. **ɦabárV* > Bsq **ha*[*m*]*aṙ* (cf. Blažek 1999).

---

[46] I have hypothesized that Bsq **-t-* in combinatory forms is a relic of a PDC oblique stem marker = PNC **-dV-*.

Bur. **haúlal > haúlal ~ ahúlal* (Y), *hoólal*-as (H, N) 'butterfly, moth'

~ Cauc: ? Archi *hiluku ~ hiliku ~ hilku* 'fly' (insect)

~ Basque **euli* 'fly' (insect).

§ Bur + Bsq compared by Berger (1956: 16), citing Zarubin's transcription of Yasin *ahúlal*. The Archi word *hil(i/u)ku* 'fly' is totally isolated within Cauc, thus dubious.

Bur. **khen > khin* (H, N), *khen* (Y) 'flea'

~ Yeniseian: PY **qəʔń* 'flea' > Ket *qʌʔń* 'flea', Kott imgara-*xon* 'flea' (*imgara-* 'little'), *xon* 'beetle'

~ Cauc: Chechen *ʙēn*-ig 'louse', Ingush *ʙon*-g id., Dargi *q'i* 'nit', Lak *q̰'unuq̰'ay*-šat:a 'worm'[47] < PEC **q̇̄ănʔV* 'louse, nit; worm' (NCED 911)

~ Basque **a=kain* 'tick'.

§ PDC **xq̇ā̆ŋʔV̆* 'louse, flea' (SCG 236). The meaning 'flea', only in Bur and Yen, may be evidence for their common development in a "Burusho-Yeniseian" subgroup of DC. (Cf. Bur **ć(h)argé* '[flying] squirrel' ~ PY **saʔqa* 'squirrel', above.)

Bur. **phirán > pherán* (Y) 'moth', *phirán* (H, N) 'spider'

~ Cauc: Chechen *polla* 'butterfly', Andi *pera* 'bee', Khwarshi *par* 'bee', etc. < PEC **pŏrV* (NCED 875); reduplicated as Andi *pirinpa* 'butterfly, Udi *päpälä*-k id., etc. < PNC **părVpăłV* (NCED 867)

~ Basque **pinpilin* > High Navarrese *pinpilin* 'butterfly', Lapurdian *pinpirin* id.

§ PDC **pŏrV́* (SCG 162).

Bur. **phen > phin* (H, N), *phen* (Y) 'fly' (insect)

~ Cauc: Avar *púq:na* 'drone', Dargi Akusha *mirqi* 'bee', Chechen, Ingush niq 'beehive', etc. < PEC **pănq̱wV* 'bee' (NCED 868)

~ ? Basque **[p]in* in (Zub) *loze-biña* 'wasp', *loza-bi* 'small wasp', (BN) *leiza-fin, leza-fin* 'wasp'.

§ PDC **pắnqwV* 'bee' (SCG 159). Presumably the Bur development was something like **pʰäŋK > *pheŋ > *phen*.

Bur. **Qarúuy̨o > qarúuy̨o* (H), *ɣarúuy̨o* (N) 'heron'

~ Yeniseian: PY [**guriraK*] > Kott *kurīrax*, pl. *kurīragan, kurīrakŋ* 'crane'

~ Cauc: Andi *q':urru* 'crane', Karata *q':uru-n,* Adyge *q:araw* id., Dargi Akusha *q'anq'* 'heron, bustard', Chechen *ʙarʙuli* 'crane', Lak *q'uruq'* id., etc. < PNC **q̱̇ərəq̇wV* (NCED 914)

~ Basque **kuṙu ~ *kuṙi ~ *kuṙ(i)-lo* 'crane'.[48]

§ PDC **xq̇ə̄́rəxq̇wV* (SCG 238), with reduplication; but some forms point to a simplex such as **xq̇ə̄́rV* (Bsq **kuṙu ~ *kuṙi*; Andi *q':urru*), or to a lateral suffix (Bsq **kuṙ(i)-lo*;

---

[47] /q'̰/ represents a pharyngealized + glottalized uvular stop/affricate.

[48] This cannot be of Romance origin, for phonetic reasons; in fact Spanish *grulla* seems to have been influenced or blended with Bsq (Bzk, Gip) *kurrillo* 'crane'. (Other Romance languages have no counterpart to the *-lla* suffix in this word.)

Chechen *ʁarʁu-li*; Bur **Qarúu-y̩o*). Bur /y̩/ is of lateral origin (see Bengtson & Blažek 2011a).

Bur. **tal* > *tal* (H, N, Y) 'dove'
~ Cauc: Avar *tł':itł':í* 'a kind of songbird', *tł':ítł':i*-diro 'a kind of variegated bird', Budukh *kak-ɨl* 'partridge', Lezgi *k'ek* 'cock', Khinalug *k'ak'-id* 'partridge', etc. < PEC **ƛ̲eƛ̣ē* 'a kind of bird' (NCED 776).
§ PDC **ƛ̣éƛ̣(w)ē* (SCG 136-137). Note recurrent correspondences of Burushaski **t*- with PDC lateral affricates (**ƛ̣*, etc.): see **Phonology** section of this paper.

Bur. **tur*- in *tur*-ćún (Y), *tur*-śún (H, N) 'marmot' (compound of **tur*- + obscure second element)
~ Cauc: Ingush *ler*-g 'hare', Andi *tł':an*-k'ala, Akhvakh *tł':ã*-k'a, Ubykh *ɮa*, Abkhaz a-*ẑá* 'hare',[49] etc. < PNC **Ł̲ărV* ~ **ƛ̣ărV* 'hare' (NCED 788).
§ PDC **ŁărHV́* ~ **ƛ̣ărHV́* (Starostin 2005d, 2007 [TOB]). Marmot and hare are both rodents. See **Phonology** section for initial correspondence.

Human relations

Bur. *=*ḉ̣u* > =*ḉ̣o* (N, H), =*ḉ̣u* (Y) 'brother (of a man) / sister (of a woman); husband of a man's sister'
~ Yeniseian: PY **b[i]s* 'brother, sister' > Ket *biśɛʔp* 'brother, sister', Yug *bisɛʔp*, Pumpokol *bič* id., Kott *popēš* 'brother', *popēča* 'sister'[50]
~ Cauc: Chechen *w=aša* 'brother' / *y=iša* 'sister', Avar *w=ac:* 'brother' / *y=ac:* 'sister', Agul *ču* 'brother' / *či* 'sister', Adyge *šə* 'brother', etc. < PNC *=*ɨ̆ć̲ɨ̆* 'brother / sister', with changing class prefixes (NCED 669)[51]
~ Basque **an-**his**-ba* 'sister (of a woman)' > (c) *ahizpa, aizpa*, (Bzk) *aizta*, (Zub) *ãhĩzpa*, (Rnc) *ãĩzpa.*

Bur. **γul* > *γul* (H, N) 'grudge, enmity, hatred'
~ Yeniseian: PY [**qər*- ~ **χər*-] > Yug *xʌrɨń*^{5} 'bad tempered, angry'
~ Cauc: Avar *ʁ*^{w}*el* 'gossip, rumor; abuse', Khinalug *qol* 'offence', Chechen *qel* '(legal) sentence, fate', etc. < PEC **ɢwāłħo* (NCED 465)
~ Basque **bilhau* ~ **bilahu* 'curse, blasphemy, hate' (< **g*^{w}*ilhau*, etc., a regular change).
§ PDC **ɢwāłħó* 'bad, angry; anger, quarrel' (SCG 55-56). An interesting 'negative emotion' etymology.

---

[49] Ubykh /ɮ/ is a voiced lateral fricative, shifted to a sibilant /ẑ/ in Abkhazian languages.

[50] Here the PDC I-class (masculine) form, corresponding to PNC **u̯=ɨ̆ć̲ɨ̆* 'brother', has been generalized for both 'brother' and 'sister'.

[51] "Two basic original forms must be reconstructed as **u̯-ɨ̆ć̲ɨ̆(jV)* 'brother' (with frequent further development > **ć̲wɨ̆jV* or **ć̲wĭjV;* exactly this form is reflected, e.g., in PL and PWC), and **j-ɨ̆ć̲ɨ̆(jV)* 'sister' (sometimes with a similar contraction > **ć̲ĭjV*)" (NCED). Thus, for example, Agul *ču* 'brother' / *či* 'sister', in which the gender element has been transferred from the former prefix to the stem vowel. In Agul the old class system is no longer grammatically productive, so the difference between *ču* 'brother' and *či* 'sister' has become purely lexical.

Bur. **hur-ik-* > *hir*, pl. *hirí* (H), *hir,* pl. *hiríkanc* (N), *hir*, pl. *hurí, huríkia* (Y) ‘man; male (of animals)’

~ Cauc: Andi, Chamali, Tindi *hek’ʷa* ‘man, person’, Inkhokwari *hik’o* ‘man, person’, Hinukh *rek’ʷe* id., etc. < PNC **Hĭrḳwɛ̆* ‘man, person’ (NCED 579); cf. Chechen *ēra* ‘ungelded’, Ingush *ärh* id., Akhwakh b=*eł:o* ‘male’, Lak b=*urx*-ni-s:a ‘male’ (the latter two with class prefixes), etc. < PEC **ʔīrλ̱wV* ‘male’ (NCED 210)

~ Basque **aṙ* ‘male’.

§ PDC **Hĭr(V)ḳwĔ́* ‘man, male’ (Starostin 2005d, 2007 [TOB]: only for Bur **hur-ik-* and PNC **Hĭr(-)ḳwĔ̆*). PEC **ʔīrλ̱wV* and Bsq **aṙ* may represent an unrelated root.

Bur. *=́s* > *=ís* (Y) ‘child; (animal’s) young’, *=́s-k* (H, N), *=́s-ko* (H) ‘(animal’s) young; (jokingly) human child’

~ Cauc: Avar w=*as* ‘son’, y=*as* ‘daughter’, Bezhta *ožo* ‘son, boy’, Lak *ars* ‘son’, Dargi *urši* ‘son’, Khinalug *ši* ‘son’, ri=*ši* ‘daughter’, Kabardian *śā*-wa ‘son’, etc. < PNC **=ĭšwE* ‘son’ / ‘daughter’, with changing class prefixes (NCED 671)

~ Basque *-*śV*- [element in kin terms], e.g. (c) *se*-me ‘son’; (Bzk) *osa*-ba ‘uncle’, gura-*so* ‘parent’, *asa*-ba ‘ancestor’; (Zub) *osá*-ba ‘uncle’, alhaba-*so* ‘grand-daughter’, *sḗmẽ-so* ‘grandson’, iloba-*so* ‘grand-nephew’, bürü-*so* ‘ancestor’, etc.

§ PDC **=ĭšwĔ* ‘son, child’ (SCG 113). Cf. Na-Dene: PA **=ya·ž(ʷ)* ‘small, woman’s child’ (Navajo *=yáázh* ‘baby [woman speaking]’, *yáásh* ‘little one’, *yázhí* ‘little, small; young [of animal]’, etc.); Eyak *yahš* ‘child (of a female)’; PST **śū* ‘grandchild’.

Bur. **ses* > *sis* (H, N), *ses* ‘persons, people’

~ Yeniseian: PY **ǯeʔ* ‘person’, pl. **ǯeʔ*-ŋ > Kott *čeäŋ* ‘people, folk’, Ket *dɛʔŋ*, Yug *d́ɛʔŋ* id.

~ Cauc: Ubykh *ćə́ća* ‘persons, people’, Avar *či* ‘man’, Tsezi *ž*-ek’u ‘man’, Lak *čuw* ‘man’, Agul *šuy* ‘man’, etc. < PNC **ćwĭjo* ‘man, male’ (NCED 336).

§ Peiros (1988, §38) compared Bur + Cauc + Yen. PDC **ćwĕjo* ‘man’ (SCG 30), reduplicated in Bur and Ubykh.

Bur. **ʒ́ām* > *jaám* (H, N), *jâm* (Y) ‘(distant) kinsman, relative’

~ Cauc: Chechen *zamō* ‘best man’, Ingush *zame* id., Rutul q’u̯-*ǯäm* ‘brother-in-law’ (*q’u̯* = ‘two’), Lezgi *č:am* ‘bridegroom’, Agul *žam* ~ *ǯam* id., Lak *mač:a* ‘kinsman’, etc. < PEC **ʒ́ămV* / **măʒ́V* (NCED 1101).

§ Peiros (1988, §63) compared Bur + Cauc. PDC **ʒ́ămV* / **măʒ́V* ‘relative’ (SCG 251).

Descriptives

Bur. *hultáṣ*, *holtáṣ* (Y) ‘barefoot’; a derivative of **=ltá-* ‘to put on (shoes, stockings) > *tá-* / *=ltá-* (H, N, Y) id.

~ Cauc: Andi *=itł’:in-* ‘to put on (shoes, footwear, trousers), Akhwakh *itł’:e*-l ‘stocking’, Tsezi *=itł-* ‘to put on (shoes)’, Archi =*ubtła*-s ‘to put on (trousers)’, etc. < PEC *=*ōmḺV* ‘to put on (trousers, shoes)’ (NCED 861)

~ Basque **ortúć* '1 barefoot; 2 to take off (shoes, stockings)' > Zub *orthǘts* 1, *orthǘs*- 2, BN *orthuts* 1, *orthus*-(tu) 2, Bzk *ortoz* 1, etc.[52]
§ Berger (1959, p. 27, note 35) compared Bur *hultáṣ, holtáṣ* + Bsq *orthuts, ortotz* 'barefoot'. PDC **ŁV́mV* 'to put on (shoes)' (SCG 130), based on a somewhat dubious PST **lomH* 'a kind of shoe',[53] altered to the Western DC verb **=ōmḺV* by a frequent process of metathesis and syllabic reduction, caused by extensive prefix- and suffixation (Starostin 2005a, p. 1). Overall an interesting DC etymology with precise phonetics and semantics. [54]

Bur. **Qaqáẏ*- > *ɣaqáẏ*-um (H, N), *qaqá*-m (Y) 'bitter; unsweetened; sour' (-*um* is a frequent adjectival ending)
~ Yeniseian: PY **qVqVr* > Ket *qʌ:ĺ; qōlɨŋ*^{1} */ qɔlaŋ*^{5} 'bitter', Yug *xʌxul*^{5}; *xɔxɨĺaŋ*^{6}, Kott *ogar* id., Pumpokol leo-*xóxar* 'bladder' ('gall-bladder'?)
~ Cauc: Chechen *q'äħa* 'bitter', Bezhta n=*iq'aro*, Archi *q'ala* id., Khinalug *q'al* 'bitter', *q'il*-ez 'salty', Ubykh *q'aq'ə́* 'sweet', etc. < PNC **q̇ĕɦlV* ~ **q̇ĕɦłV* 'bitter' (NCED 912)
~ Basque **keru* 'stench; rancor'; **kerać* 'bitter; foul-smelling'.
§ Bur and Basque compared by Berger (1956: 10). PDC **xq̇ĕɦlV̄́* ~ **xq̇ĕɦłV̄́* 'bitter' (SCG 236-237). The stem is reduplicated in Bur, Yen, and West Cauc.

Bur. **ṣuq*-úr- > *ṣuq*-úr-um (H, N) 'sour, bitter', =*ṣq*-ur- 'to sour, turn sour' (H, N), *iṣq*-ór-um (Y) 'sour, bitter', í=*ṣq*-ur (H) 'sourness'
~ Cauc: Chamali *s'ik'u*- 'sour', Khwarshi *caqu*, Lak *c:iχku-*, Archi *c'eg*^{w}-du 'rank, bitter', etc. < PEC **c̱ā̆ḳwV* 'sour, raw' (NCED 356)
~ Basque: (with metathesis **c̣ā̆xḳwV* > **xḳwā̆c̣V*) **gasi* 'salty', (dial.) 'acidic; bitter', **gastana* 'cheese', **gac* 'salt'.
§ PDC *c̣ā̆xḳwV* 'sour; bitter' (SCG 24), with metathetic variant **xḳwā̆c̣V* to account for Bsq **gac / *gas- / *gast-*. Cf. PST **sǎk* 'bitter, pungent'; Na-Dene: Eyak *c'iʔk*' 'bitter'; PA **-č'i·k'* 'to sting, smart; be peppery, bitter' (Leer 1993).

Bur. **tharén- > tharén*-um (H, N) 'narrow, cramped, tight (of clothes)'
~ Cauc: Avar *t'eréna-* 'thin', Karata *=etł'ara-*, Dargi Akusha b=*uk'ula*, Khinalug *k'ɨr*, id., etc. < PNC **=iƛ̣ɨlV* 'thin' (NCED 639)
~ Basque **lirain* 'slender, svelte, lithe, agile, graceful'.
§ Bouda (1950, §114) compared Bur + Avar; Berger (1959, p. 26) compared Bur + Bsq. PDC **=iƛ̣ə̆lV́* 'thin' (SCG 105). For Bur aspirated initial (for expected **t*-) cf. Bur **thugár* 'buck goat', below.

Cultural vocabulary: domestic animals

---

[52] Basque **-rt-* is the normal correspondence to Bur **-lt-* and PNC **-ƛ-, *-ƛ̣-, *-Ł-*, when the following vowel in PDC is accented, e.g. Bsq **urte* 'year' < PDC **ʔVƛ̣wV́* 'last year' (SCG 259).

[53] Based only on Tibetan *lham* 'boot, shoe', Old Chinese 踊 **loŋʔ* 'shoes for criminals with lopped toes' ('shoes [with lopped toes] for criminals'?).

[54] It is interesting that this verb is also found in Na-Dene: e.g. Navajo *=tłé, =tłèèˀ*, in *yìstłé, =stłèèˀ* 'socks, stockings, leggings'.

Bur. *_aćás_ > _aćás_ (H, N, Y) ‘sheep, goat, sheep and/or goat(s) = _Kleinvieh_, small cattle’
~ Cauc: Adyge _āča_ ‘buck goat’, Dargi Akusha _ʕeža_ ‘goat’, Chechen _awst_ ‘goat (about 1 yr. old)’, etc. < PNC *_ʔēyʒ́wē_ (NCED 245).
§ PDC *_ʔēy[ʒ́]wḗ_ ‘goat’ (SCG 264). Note, however, similarity, even in accent, with Old Indic _ajá-_ ‘goat’ (PIE *_aǵ-_), raising the possibility of borrowing, but this word is known only in eastern (_Satəm_) IE languages. The frequent semantic variation ‘sheep’ ~ ‘goat’ recurs in several of the comparisons below.

Bur. *_buć_ > _buć_ (H, N) ‘(ungelded) male goat, 2 or 3 years old’
~ Cauc: Lak _buχca (< *buc-χa?)_ ‘buck goat (1 year old)’, Rutul _bac’i_ ‘small sheep’, Lezgi _bac’i_ ‘kid’, Khinalug _bac’iz_ ‘kid’, etc. < PEC *_b[a]c̣V_ (NCED 287).
§ Bouda (1950, §139) compared Bur + Lezgi, etc. Note variation between the meanings ‘young goat’ and ‘young kid’ within the East Caucasian family. Berger (1998: 60) notes a similar word in (Iranian) Wakhi, _buč_ (< Burushaski?). But there are also Avestan _būza-_ ‘buck goat’, Persian _buz_ ‘goat’, etc. PDC *_bV[c̣]V_ ‘kid, goat’ (SCG 17).

Bur. *_chigír_ > _cigír_ (Y), _chigír_ (N), _chiír_ (H) ‘(nanny-)goat’, also ‘female ibex’
~ Cauc: Karata _c’:ik’er_ ‘kid’, Godoberi _c’ek’ir_, Tsezi _cek’i_ ‘kid’, Lak _c’uku_ ‘goat’, Adyghe _č́ac’ə_ ‘kid’, etc. < PNC *_ʒĭkV̆_ / *_k̲ĭʒV̆_ ‘goat, kid’ (NCED 1094)
~ Basque *_siki-ro_ ‘gelded ram’, *_siki-te_ ‘gelded goat’.
§ PDC *_sdĭ́[k]V̆_ ‘goat’ (SCG 187).[55] Note the similar suffixes in Bur *_chigí-**r**_, Andian *_c̣:iḳi-**r**_ / *_c̣iḳi-**r**_ and Bsq *_siki-**ro**_.

Bur. *_chul-_ > _chulá_ (H, N), _culá_ (Y) ‘male breeding stock’: (H) ‘drake’, (N, Y) ‘buck goat’; _culdár_ (Y) ‘bull’, _chindár_ (H, N) id.
~ Cauc: Andi _č’ora_ ‘heifer’, Tindi _č’ara,_ Agul _luč’_ id., Chechen _ēsa_ ‘calf’, Ingush _ʕasa_ id., etc. < PEC *_Hć̣wɨ̄lV̄_ / *_Hlɨ̄ć̣wV̄_ ‘heifer’ (NCED 556)
~ Basque *_čahal_ > (Zub) _xáhal_ /šáhal/ ‘calf, heifer’, (Bzk) _txaal_ /čaal/ ‘calf’, etc.
§ PDC *_H[ć̣]wɨ̄lV̄́_ ‘bull, heifer’ (Starostin 2005d, 2007 [TOB]); with metathesis *_[ć̣]VHɨ̄lV̄́_ or *_[ć̣]ɨ̄HVlV̄́_ to account for Bsq *_čahal._

Bur. *_du[m]_ > _du_ (H, N, Y), _dúdo_ (H) ‘kid, young goat up to one year’ (< *_dũ_)[56]
~ Cauc: Andi _dan_ ‘sheep, ewe’, Chechen _tō_ ‘ram’, Lak _t:a_ ‘sheep, ewe’, Kabardian _t’ə_ ‘ram’, etc. < PNC *_dwănʔV_ ~ *_dwănhV_ ‘sheep, ram’ (NCED 405)
~ Basque *_i=di_ > (c) _idi_ ‘bull’.[57]

---

[55] PDC *_sd_ has reflexes similar to PDC *_ʒ_ [d^z] in Bsq (*_s_) and PNC (*_ʒ_), but the reflexes differ in Bur and PY: PDC *_sd_ > Bur *_ch_, PY *_t_, but PDC (*_ʒ_) > Bur *_s_, PY “*_ʒ_ or *_s_, with distribution yet unclear” (SCP, pp. 53, 75-76). Since the Yeniseians did not practice animal husbandry (apart from some reindeer herding at a late date) and cereal cultivation, PY lacks cognates for many of the cultural words discussed here.

[56] Loss of PDC nasals (via nasalized vowels) is frequent in Bur (cf. convergent cases in Cauc and Bsq), with evidence for the original nasal often preserved in inflected forms. In this case *_du[m]_ is hypothesized based on (H) _dúdo_ ‘kid’, pl. _dúdomuc._

[57] For semantic typology, cf. Welsh _dafad_ ‘sheep, ewe’ ~ OIr _dam_ ‘ox’, Gk δαμάλης ‘young ox’ (< ‘tamed animal’: Buck 3.25). The loss of PDC nasals (in clusters of the type *_-nH-_) is also frequent in Bsq.

§ A different comparison of Bur **du[m]* (with PNC **s̱wänʔV* 'lamb') was preferred by Starostin (SCG 191).

Bur. **ḍágar* > *ḍágar* (N) 'ram'
~ Cauc: Avar *deʕén* 'billy-goat, buck goat', Hinukh *t'eq'ʷi* 'kid (about 1 year old)', Inkhokwari *t'ɨq'o*, Bezhta *t'öq'ä* id., etc. < PEC **dVrq̇wV* 'billy-goat' (NCED 403).
§ PDC **dVrxq̇wV́* 'male animal' (SCG 43). Initial retroflex /ḍ/ in Bur apparently conditioned by following /r/.

Bur. **hálgi-t* > *élgit* (H, N), *hálkit* (Y) '(female) goat, over 1 year old, which has not given birth'[58]
~ Cauc: Agul, Tsakhur *urg* 'lamb (less than a year old)', Rutul *urg* 'yearling sheep', Chamali *bargʷ* 'a spring-time lamb', etc. < PEC **ʔwɨlgɨ* (NCED 232).
§ PDC **ʔwəlgɨ́* 'lamb, kid' (SCG 260). Perhaps **ʔəlgɨ́* is better, since the PEC form may incorporate a class prefix: **w=ʔəlgɨ́* > **ʔwɨlgɨ*, not uncommon in PNC/PEC forms.

Bur. **huyés* > *huyés* (H, N, Y) '*Kleinvieh*, small cattle, sheep and/or goats'
~ Cauc: Avar *ʕi* 'flock (of sheep)', Lak *ya*-t:u 'flock (of sheep)' < PEC **ɦVʔV̆* ~ **ɦVNHV̆* 'flock' (NCED 532).[59]
§ PDC **ɦVʔV̆* 'sheep, small cattle' (Starostin 2005d, 2007 [TOB]); however there is a note that Bur **huyés* "should be moved to [PDC]**χVjV!*," i.e. **χVHV́* 'a kind of deer' (PST **ɣŭ* 'doe, antelope'; PY **χV̄j* 'elk, deer'), which seems less plausible to me. Cf. also PIE **Howi-* 'sheep'.

Bur. **thugár* > *thugár* (H, N) 'buck goat, billy-goat'
~ Cauc: Karata *t'uka* 'buck goat', Tindi *t'uḱa* ~ *ḱ'uta* (in free variation), Bezhta *t'iga,* Kabardian *daǯa-χʷ* 'buck goat', etc. < PNC **ṭūgV̄* (NCED 1003).
§ PDC **[ṭ]ūgV̄́* 'goat' (Starostin 2005d, 2007 [TOB]). The Bur initial **t-* (unaspirated) would be expected. The aspirate seems to be the result of a kind of "Verner's Law" effect, caused by accent on the following syllable (cf. Bur **tharén-* 'narrow', above).

Cultural vocabulary: dairy[60]

Bur. **ćham-* > *ćhamaní* (H) 'food rich in butter', *ćamaní* (Y) 'food rich in butter', *ćámu* (*mamú*) 'beestings'
~ Cauc: Lezgi *č'em* 'butter', Agul *č:ạm* id., Hinukh *cen* 'curds', Khinalug *mic* 'butter, oil', etc. < PEC **Hʒ́ĕmɨ̄* 'butter, curds' (NCED 624).

---

[58] TOB adds note: "Could be derived from **halk-* `to bear young' (not quite plausible semantically, however); or could be secondarily contaminated with this root in Yas[in]."

[59] "An exclusive Avar-Lak isogloss; very unsecure in what concerns PEC reconstruction" (NCED).

[60] "Evidence for dairying activities in prehistory can be assessed by the detection of dairy fats associated with archaeological pottery … This has revealed that milk was being processed in the northwestern part of present-day Turkey … as early as 8500 years … BP" (Gerbault, et al. 2013). This is well within the postulated time range of the Western DC family (see **Postscript**).

~ Basque: **sen-[-bera]* 'soft cheese, cottage cheese' > Salazarese *zenbera*, etc.
§ PDC **H[ʒ́]ĕ́mī* 'butter, tasty food' (Starostin 2005d, 2007 [TOB]).

Bur. **c̣háo* > *c̣háo* (+ verb =*t́*- 'do, make') (H, N) 'to milk'
~ Cauc: Lezgi *ac:a*- 'to milk', Archi =*ac:a*-, Dargi Akusha =*iz*-es, Lak t:-*izi*-n id., Ubykh *ʒwa*- 'to drink' (< '*to drink milk' < 'to milk'), etc. < PNC *=*āmʒŬ* 'to milk' (NCED 262).
~ Basque **e=aic̣i* 'to milk (a cow)' > Zub *jaitzi*, Goizueta (Navarre) *jetzi*, etc.
§ PDC *=*āmśdŬ́* 'to milk' (SCG 4). There are several nominal derivatives, such as Tindi, Karata *zini* 'cow' = Basque **sesen* 'bull'.

Bur. **dílta-r* > *díltar* 'buttermilk' (H, N, Y)
~ Cauc: Hunzib *reł* 'butter', Tsezi *ril* , Khwarshi *lał* id., Avar *rax* 'milk', etc. < PNC **rħăƛwV̆* 'milk' (NCED 949).
§ PDC **ŕħăƛwV̆́* (SCG 183-184). For the regular correspondence of Burushaski *-lt- (-ld-)* to Cauc *-*ƛ*- [tɬ], see the **Phonology** section of this paper. PDC **r*- > Bur **d*- is also regular (SCP 41).

Bur. **malt-áṣ* > *maltáṣ* (H, N, Y) 'butter'
~ Cauc: Archi *nạtł':* 'milk', Udi *naq:* 'buttermilk', Chechen *nalχa* 'butter', Lak *nak'* 'milk', etc. < PEC **nħĕƛ̱V* 'milk, dairy product' (NCED 849).
§ PDC *[*m*]*ħĕ́ƛ̣V* 'milk, butter' (SCG 146). The **n*- in Cauc is still unexplained: possibly by contamination with PNC **rĕnχwĂ* 'butter' (NCED 948), which has *n*- initials in several languages (Avar *naχ* 'butter, oil, fat', Lak *nah* 'butter', Dargi Akusha *nerχ* 'melted butter', etc.).

Bur. **ṣiŋ* > *ṣiŋ* (H, N) 'milk, yield of milk' ('Milch, Milchertrag')
~ Yeniseian: PY **de(ʔ)n* 'milk, nipple' > Pumpokol *den* 'milk', Arin *téŋ*-ul 'milk', Kott *ten* 'nipple' (with regular change of PDC **š*- > PY **d*-: see SCP 63)
~ Cauc: Chamali *s:ī̃w* 'milk', Godoberi *š:ĩwu*, Botlikh *š:ĩʔu, š:iʔũ* id., Chechen *šin* 'udder', Ubykh *č́a* 'milk', etc. < PNC **s̱̆ămʔV* 'milk, udder' (NCED 982)
~ Basque **e=śene* 'milk' > *esene, esne, ezne*; **sen*-(bera) 'soft cheese' is a distinct etymon (see Bur **ćham-*).
§ PDC **šăŋ́ʔwV* 'milk, nipple' (SCG 195-196).

Cultural vocabulary: the horse

This topic is rather difficult. Archeological evidence indicates that horses were domesticated on the steppes of Central Asia (now Kazakhstan and Russia), certainly by 3500 BCE and possibly as early as 4500 BCE (Outram, et al. 2008; Anthony & Brown 2011). However, according to recent glottochronological results of the Moscow School the "break-up of the North Caucasian–Basque and Yenisseian–Burushaski branches [of Western Dene-Caucasian took place in] the second half of the 9th millennium BC" (Kassian 2010a: 430), thus about four millennia before domestication of the horse. Of course, the horse existed as a hunted wild animal long before that. In the 2001 version of this paper and others I have suggested the comparison of Bur **hayúr* 'horse' with PNC **farnē* 'horse' (Adyge *fāra*, Kabardian *xʷāra* 'thoroughbred horse', Lezgi *χʷar* 'mare', Khwarshi *χaram* 'foal', etc.: NCED 425) and Basque **behoṙ* 'mare', but the phonetics involved in this are quite problematic. Berger (1998: 185) notes the resemblance of Bur **hayúr* to Turkish *aiğır* 'stallion', which seems to be a plausible source, since the Turks have a long history of mounted nomadism in central Asia.

However, the words for 'saddle' in Caucasian and Burushaski are very similar and fit the correspondence of Bur **t- / *-lt-* to PNC **ƛ̣* (and other lateral affricates) that recurs in several etymologies discussed here (see **Phonology**).

Bur. **ltul[i] > =ltúl* (H, N, Y) 'to saddle (a horse), prepare mount', *tilíaŋ* (H, N), *tilíhaŋ, teléhaŋ* (Y) 'saddle' (noun, with frequent plural morpheme *-aŋ*)
~ Cauc: Avar *tł':ilí* 'saddle', Akhwakh (with assimilation) *tł':etł':e,* Andi *tł':iru,* Lak *k'ili,* Dargi Akusha *gili*, Sirgokala, Tsudakhar *guli* 'saddle', etc. < PEC **ƛ̱wiłē* 'saddle' (NCED 783).
§ Peiros (1988, §65) compared Bur + Cauc.

Of course, the phonetic correspondences do not necessarily prove that the Bur and Cauc words are "genetic" cognates; they could be early loanwords in either direction, or from an unknown source, which I have not been able to track down. Note another interesting comparison involving equids:

Bur. **ćhardV́ > ćhardá* (H, N), *ćardé* (Y) 'stallion'
~ Cauc: Abkhaz a-*čada, čadə* 'ass, donkey', Adyge *š́əd*, Kabardian *šəd*, Ubykh *čədə* id. (Čirikba 1996: 314)
~ Basque **a=śto* or **ar=śto* 'donkey, ass' > (c) *asto*, (Zub-archaic) *arsto* 'donkey, ass' (with usual syncopation of *-*śVto* > *-*śto*).

The domestication of the donkey is also much later than the postulated breakup of Western DC (Marshall & Weissbrod 2011), but the possibility remains of a designation of the wild ass or other equid. Berger (1959: 32, note 55) pointed out the unlikely similarity of Bur *j̣akún* 'ass' (with retroflex initial = [ḍẓ̌]) and Hausa *žāki* 'ass', pl. *žakūna*, which, if genuine, would have to be traced to trade contacts.

Cultural vocabulary: cereal cultivation and processing

Bur **baỵ > baỵ* (H: double plural *baçéŋ*), *baỵ* (N: double plural *báỵiŋ*), *ba* (Y) '(small-grained) millet' (Panicum miliaceum)
~ Cauc: Chechen *borc* 'millet', Karata *boča,* Avar *muč* (<**binčʷa*)Adyge *mašə* 'millet', etc. < PNC **bŏlćwĭ* (NCED 309).
§ PDC **bŏlćwĭ́* 'millet, rice' (SCG 15: also including PST **phrē(s)* 'rice'). Bur **ỵ* often derives from PDC laterals or lateral clusters (see Bengtson & Blažek 2011).

Bur. **ćha > ćha* (H, N), *ća* (Y) 'millet' (Setaria italica)

~ Cauc: Bezhta _č'e_ 'a species of barley', Andi _č'or_ 'rye', Kryz _č'eʕ_ 'roasted grain', etc. < PEC *_č̣[e]ħlV_ 'a kind of cereal' (NCED 384).
§ The loss of PDC *_-ħl-_ in Bur is difficult to verify. In NCED this cluster is only found in this entry and *_kăħlV_ 'a kind of berry', the latter with no known Bur cognate. Starostin (2005d, 2007) preferred a different comparison, with PST *_Ćɨă_ 'hemp', but also citing Proto-Nakh *_ča_ 'straw' (cf. NCED 978) and Budukh _čeʕ_ 'roasted grain' (cf. Kryz _č'eʕ_ id., with glottal initial, above).

Bur. *_daltán-_ > _daltán-_ (N) 'to thresh (millet, buckwheat)' (< *_rVŁV́-n-_)
~ Cauc: Ingush _ard-,_ Batsbi _arl-_ 'to thresh', Tindi _=elĩ-_ 'to thresh, _rali_ 'grain ready for threshing', Bezhta _=ol-_ 'to thresh', Archi _tłorom_ 'threshing board', etc. < PEC *_=V̄rŁV_ 'to thresh', *_r=ĕŁō_ 'grain ready for threshing' (NCED 1031)
~ Basque *_laṙain_ 'threshing floor'.
§ PDC *_rVŁV́_ 'to thresh' (SCG 182) ~ *_ŁVrV_ (> Bsq).

Bur. *_darc_ > _darc_ 'threshing floor, grain ready for threshing'
~ Cauc: Lak _t:arac'a_-lu 'threshing floor', Dargi _daraz_, Tabasaran _rac:,_ Lezgi _rat_, Andi _hinc':u_, Bezhta _ãc_ id., etc. < PEC *_ħrə̄nʒ̱ū_ (NCED 503).
§ Bouda (1950, §4; 1954, p. 228, §4) compared Bur + Lak. PDC *_ħrə̄́nsdū_ (SCG 100). Bur development is similar to those of eastern Dagestan (Lak, Darg, Lezg), with *_r_ > *_d_ (SCP 41).

Bur. *_gur_ > _gur_ (H, N, Y) 'wheat', _gur-gán_ (H, N) 'wheat sown in autumn'
~ Cauc: Agul _q'ir_ (dial. _q'ur_) 'grain', Rutul _q'ir_ 'winter wheat', Udi _ar_-um 'wheat' (< *_ʔar-_ < *_q'ar-_), Dargi _q'ar_ 'grass', etc. < PEC *_q̇[ə̄]rV_ 'a kind of weed, (wild) cereal' (NCED 915)
~ Basque *_gaṙ-_ in *_(gara-)gaṙ_ 'barley': a compound with *_ga[l]i_ 'wheat'.[61]
§ Berger (1998: 161) notes the similar Tibetan word, _gro_ 'wheat'. Starostin (SCG 243) compares instead Tibetan _khre_ 'millet', etc. < PST *_khriə̆H_ 'a kind of grain'.

Bur. *_harṣ_ > _harṣ_ (H, N), _harṣ, haṣç_ (Y) 'plow'
~ Cauc: Akhwakh _ʕerc:e_ 'wooden plow', Godoberi _rec:i_, Lak qa-_ras_ id., Chechen _ästa_ 'hoe, mattock; plane (for woodworking)', etc. < PNC *_Hrāyc̱ū_ 'wooden plow; mattock' (NCED 601)
~ Basque *_hai(n)cu-ṙ_ > Basque (Lab) _haintzur_ 'hoe', (Rnc) _aintzur_ '(heavy) hoe', (Zub) _háitzür_ 'pickax', (Bzk) _atxur_ 'spade', (Gip) _aitzur_ id., etc.
§ PDC *_Hrā́y[c]ū_ 'plow' (Starostin 2005d, 2007).

Cultural vocabulary: other artifacts:

---

[61] In the 2001 version of this paper I compared Bur *_gur_ 'wheat' with PEC *_ɢ̱ōlʔe_ 'wheat' and Bsq _gari_ (*_gali_) 'wheat', but now the SCG comparison seems preferable phonetically.

Bur. **baltí* > *baltí* (Y) 'front room of house', *baldí* (H, N) 'veranda'

~ Cauc: Hinukh *butłe* 'house', Hunzib *butłi* 'at home', Lak *burč'a*-lu 'threshold', Archi *notł':* 'house, room', etc. < PEC **bŭlƛ̱V* 'house' (NCED 312).

§ PDC **bŏ́lƛ̣V* 'house' (SCG 15). With the frequent correspondence of Bur *-*lt*- ~ PDC *-*ƛ̣*- (see **Phonology**).

Bur. **cháɣur* > *cháɣur* (H, N) 'chest or box for grain or meal'

Cauc: Avar *caʁúr* 'corn bin, barn', Chechen *cχar* 'penthouse', Khinalug *cuqa* 'shed, cattle-shed' < PEC **cVGVrV̌* 'shed, penthouse' (NCED 328).

§ Bouda (1950, §121) compared Bur + Avar, a resemblance also noted by Berger (1998 III: 73).

Bur. **khor* > *khor* (H, N) 'large wicker basket'

~ Cauc: Akhwakh *koro* 'trough, gutter', Andi *koru* 'a kind of jar', Lezgi *k:*^{w}*ar* 'a big jar (for carrying water)', Tabasaran, Agul *g*^{w}*ar* id., etc. < PEC **ḵwərV* (NCED 706).

§ PDC **kwə́rV* 'a kind of vessel, scoop' (SCG 118).

Bur. **tókur* > *tókur* (H, N) 'wooden chest (for grain, etc.)'

~ Cauc: Ingush *t'aqa* 'tub (for cheese brine)', Dargi Akusha *t'aq'a* 'hive', Agul *t'ak* id., Rutul *t'ak* 'basket (for berries)', etc. < PNC **ṭāqV* 'a kind of vessel' (NCED 997).

§ SCG (200) has a different comparison, with PEC **tāḳwV* 'a kind of vessel', which is more suitable for the second consonant (Bur **k* = PEC **ḳ*), while the above comparison is more suitable for the first consonant (Bur **t* = PNC **ṭ*), with secondary change of PDC **q* > Bur **k* (as in Lezgian). Possibly further study will determine which option is best.

Bur. **(y)alt*- in *giy̯ált* (H, N) 'spoon, scoop' (compound with **giy̯*- 'to pour, put in, put on', etc.)

~ Cauc: Hunzib *ẽgu* 'wooden shovel', Khwarshi *ãku*, Lezgi *yirf* id., etc. < PEC **yă[l]ƛwV* 'wooden shovel' (NCED 673)

~ Basque **śa-ɦarde* 'rake; (two-tined) fork; (dinner) fork': compound with obscure first element **śa*- (perhaps a haplologic compound with Bsq **śarha-tu* 'to clear land', i.e. **śarha-ɦarde* > **śáɦarde*?).

§ PDC **yălƛwV* 'shovel, ladle' (SCG 113). With the frequent correspondence of Bur *-*lt*- ~ PDC *-*ƛ*- (see **Phonology**).

Bur. **yeéś* > =*yeéś* (H, N) 'lasting dwelling place, permanent residence'

~ Cauc: Tsakhur *yic'a* 'sty, cattle shed', Chechen *c'a* 'house, room', Khinalug *c'*^{w}*a* 'house', Ubykh *c*^{w}*əyá* 'house, room', etc. < PNC **c̣[ī]yu* ~ **c̣ūyV* ~ **y[ī]c̣u* 'house' (NCED 364)

~ Basque **eće* 'house' (or **e=će*, with fossilized class prefix = PNC **i̯*=) > (Bzk, Rnc) *etse*, (c) *etxe* /eče/ 'house'. The common form *etxe* seems to have been the expressive form, now generalized.

§ Bur + Bsq compared by Berger (1956: pp. 18, 24). PDC *c̣[ī]jú* 'house' (Starostin 2005d, 2007).

Phonological Correspondences

Correspondences of stops: In general, Burush aski unaspirated **p, *t/ṭ, *k, *q* correspond to PNC/PEC glottalized **ṗ, *ṭ, *ḳ, *q̇* (= [p', t', k', q']), respectively; and Burushaski aspirated **ph, *th, *kh, *qh* correspond to PNC/PEC aspirated **p, *t, *k, *q* (= [p^h, t^h, k^h, q^h]), respectively. For voiced stops the correspondences are generally trivial, i.e. **b* = **b, *d/ḍ* = **d, *g* = **g, *ɣ* = **ɢ*.[62] See the following examples from the lexical comparisons cited above (some from S.A. Starostin's later etymologies):

Bur. **p, *t/ṭ, *k, *q* = PNC/PEC **ṗ, *ṭ, *ḳ, *q̇*

- Bur. **tápi* 'stony terrace' ~ PEC **ƛ̣ĕṗV̄* 'stone plate or shed' (SCG 137)
- Bur. **ltopo, *tultopo* 'a kind of bread' ~ PEC **HārŁ̱āṗV* 'a kind of food' (SCG 63)
- Bur. **tókur* 'wooden chest' ~ PNC **ṭāqV* 'vessel': Rutul *t'ak* 'basket (for berries)'[63]
- Bur. **=ḥuṭ* 'foot' ~ PEC **ħīṭwV̄* ~ **ṭwīħV̄* 'foot': Avar *ħet'é / ħet'*
- Bur. **=ḱen* 'liver' ~ PEC **ḵunHV* 'kidney': Chamali *k'ũ̄* 'liver'
- Bur. **kur-* 'bone, joint' ~ PEC **ḳwĭrV* 'bone, leg': Lezgi *k'ur* 'leg, hoof'
- Bur. **tumá-ỵ* 'nut, fruit' ~ PNC **ṭŭmhV* 'kernel, fruit': Archi *t'ummu*-l 'grape'
- Bur. **túni* 'small basket' ~ PEC **ṭwōnʔe*: Dargi *t'eni* ~ *t'uni* 'trough'

Bur **ph, *th/ṭh, *kh, *qh* = PNC/PEC **p, *t, *k, *q*

- Bur. **phen* 'fly' ~ PEC **pănq̱wV* 'bee': Avar *púq:na* 'drone'
- Bur. **phirán* 'moth, spider' ~ PEC **pŏrV* 'bee': Tindi *pera* 'bee'
- Bur. **phunc* 'dew' ~ PNC **pĭnc̱wĂ* 'resin, juice': Lak *pic'* 'dew; sweat'
- Bur. **ther* 'dirt' ~ PEC **türV* 'dirt, dung': Akhvakh *tere*-ti 'ashes, dust'
- Bur. **khor* 'large basket' ~ PEC **ḵwərV* 'vessel': Akhwakh *koro* 'trough, gutter'
- Bur. **=qhaś-* 'arse, genital' ~ PNC **=Vq̱V* 'behind': Udi *qoš*
- Bur. **=qhát* 'mouth' ~ PEC **qwīṭi*: Lak *qiṭ'* 'Adam's apple; beak'
- Bur. **=qhorpVt* 'lung' ~ PEC **q̱wəlθV(rV̄)*: Archi *χurt:ur*-t:i 'lung'

Bur. **b, *d/ḍ, *g, *ɣ* = PNC/PEC **b, *d, *g, *ɢ*

- Bur. **bác-in* 'shank (of animal)' ~ PEC **b[ə]c̱V*: Chamali *bec*^{w} 'knee (of animal), thigh'
- Bur. **baltí* 'front room, veranda' ~ PEC **bŭlƛ̱V* 'house': Hunzib *butłi* 'at home'
- Bur. **baŋ* 'resin' ~ PEC **bħĭnḵwV*: Chechen *baga* 'pine tree'
- Bur. **baỵ* 'millet' ~ PNC **bŏlćwĭ*: Tindi *boča* 'millet'
- Bur. **buć* 'goat' ~ PEC **b[a]çV* 'goat': Lak *buχca* 'yearling male goat'
- Bur. **búmbal-* 'ankle' ~ PEC **bĭ̄młV̆*: Tsezi *bula* 'hoof'; cf. PY **bul* 'foot, leg'
- Bur. **dul-dúm* 'cloud of dust', etc. ~ PEC **dilƛ̱wV*: Lak *t:urlu* 'cloud'
- Bur. **du[m]* 'kid' ~ PNC **dwănʔV* 'sheep, ram': Lak *t:a* 'sheep, ewe'
- Bur. **ḍágar* 'ram' ~ PEC **dVrq̇wV* 'billy-goat': Avar *deʕén* id.
- Bur. **ćhardV́* 'stallion' ~ Abkhaz a-*čada, čadə* 'ass, donkey', etc.

---

[62] Retroflex stops have sometimes developed secondarily in Bur from original dental stops. The conditioning factors are not yet fully understood. See Bengtson & Blažek (2011: 3-4).

[63] PEC has another similar root, **tāḳwV* 'a k. of vessel' (NCED 990). See the note under **ṭāqV* (NCED 997). The initial **t-* of Bur. **tókur* fits best with PNC **ṭāqV*, but the internal **-k-* fits **tāḳwV*.

- Bur. **gal* ‘cross-beams’, etc. ~ PEC **gălV* ‘stick’: Khwarshi *gelu* ‘pole’ (SCG 50)
- Bur. **gaṭú* ‘clothes’ ~ PEC **gwĭrdwV*: Avar *gordé* ‘shirt’
- Bur. **ć(h)argé* ‘flying squirrel’ ~ PNC **cārg̱wV* ‘weasel, mouse’
- Bur. **hálgi-t* ‘(female) goat’ ~ PEC **ʔwɨlgɨ* ‘lamb’
- Bur. **=ɣā́n* ‘heel’ ~ PEC **ʔĭnɢwV̆* ‘heel’, etc. < PDC **ʔĭɢwVnV́*
- Bur. **ɣul* ‘hatred’, etc. ~ PEC **ɢwāłħo* ‘offence’, etc.
- Bur. **cháɣur* ‘chest for grain’, etc. ~ PEC **cVɢVrV̆* ‘shed, penthouse’

There remain a few seemingly anomalous cases, but there is a growing body of evidence that some reflexes have been conditioned by the position of accent, for example, a tendency for unaspirated stops to become aspirated (or, in the case of uvulars, **q > qh ~ ɣ*) when the accent is on a following (or preceding) syllable. (Cf. “Verner’s Law” in Indo-European phonology.) Some dialectal variations in these tendencies (*q ~ qh ~ ɣ*) can be observed:

- Bur. **phol-(ɣúuỵ)* ‘feather’ (expected **pol-*) ~ PNC **ṗVħVłV* ‘feather’, etc.
- Bur. **tharén-* ‘narrow’ (expected **tar-*) ~ PNC **=iƛ̣ɨ̄lV* ‘thin’
- Bur. **thugár* ‘buck goat’ (expected **tug-*) ~ PNC **ṭūgV̄* ‘buck goat’
- Bur. (H) *qarúuỵo* ~ (N) *ɣarúuỵo* ‘heron’ ~ PNC **q̇̄ərə̄q̇wV* ‘crane’
- Bur. (Y) *qaqá*-m ~ (H, N) *ɣaqáỵ*-um ‘bitter; unsweetened; sour’ ~ PNC **q̇ĕɦlV* ~ **q̇ĕɦłV* ‘bitter’
- Bur. (H) *=́qat* ~ (N) *=́qhat* ~ (Y) *=́qet*-araŋ ‘armpit’ ~ PEC **qVdV* ‘brisket’ (here the accent is on a preceding pronominal prefix)[64]

Correspondences to Caucasian lateral affricates: As already pointed out in some of the notes to the above lexical comparisons, there are recurrent correspondences between Burushaski initial **t-* and medial **-lt-* and the Proto-Caucasian lateral affricates **ƛ*, **ƛ̣*, **Ł* = [**t^l, t^{l}’, d^l**], as shown in the following examples. The corresponding reflexes in Basque are initial **l-* and medial **-rd-*, respectively. (Comparisons already listed above will be cited in abbreviated form):

(a) Examples with Burushaski initial ****t-***:

- Bur. **taɣ* ‘branch, shoot’ ~ PEC **ƛ̣ōrχwV* ‘sprout’
- Bur. **tal* ‘dove’ ~ PEC **ƛ̱eƛ̣ē* ‘a kind of bird’
- Bur. **tal* ‘stomach, belly’~ PEC **Hƛ̣älV* ~ **Hläƛ̣V* ‘liver’
- Bur. **táno* ‘colon, rectum’ ~ PNC **Hƛ̣ŏnŭ* ‘bottom’
- Bur. **tápi* ‘stone terrace’ ~ PEC **ƛ̣ĕṗV̄* ‘stone plate, shed’ ~ Bsq **lape* ‘shelter’
- Bur. **tar[í]-* ‘*skin’ ~ PNC **Ł̱ŏli* ‘*skin’ ~ Bsq **laṙu* ‘skin, hide, leather’
- Bur. **ter* ‘summer pasture’ ~ PEC **ŁwĕłV* ‘enclosure, fence’ ~ Bsq **laṙe* ‘grassland, pasture’
- Bur. **tiṣ* ‘wind’ ~ PEC **ƛ[a]rʒ́V* ‘movement of air’
- Bur. **tur*-Cún ‘marmot’ ~ PNC **Ł̱ărV* ~ **ƛ̣ărV* ‘hare’
- Bur. **tharén*-um ‘narrow, tight’ ~ PNC **=iƛ̣ɨ̄lV* ‘thin’ ~ Bsq **lirain* ‘lithe, svelte’

(b) Examples with Burushaski alternation of (initial) ****t-*** / (medial) ****-lt-***:

---

[64] Realized as (H) *á=qat* ‘my armpit, *gó=qat* ‘thy -‘, *mó=qat* ‘her -‘, etc.‘

- Bur. **ltén* (Y) *ten* 'bone' / (H, N) *=ltín* 'bone' (bound form) ~ PEC **ƛ̲wVnʔV* 'ankle, shin'
- Bur. **ltur>* (Y) *tur* 'horn' / (H, N) *=ltúr* (bound form) ~ PEC **ƛ̲wɨ̆rV* 'horn; braid, mane' ~ Bsq **a=daṙ* 'horn' (< **a=rdaṙ*)
- Bur. **ltap* > *tap* 'leaf' / du=*ltápi-* ~ *=ltápu-)* 'to wither' ~ PNC **ƛ̣ăpi* 'leaf' ~ Bsq **lapa-ṙ* 'bramble'
- Bur. **lté* > (Y) *té-* / *=lté-* 'to swear' ~ PEC **HiŁ̲V* 'to say' (NCED 572)
- Bur. **ltul[i]* 'to saddle', **tilí-* ~ **telé-* 'saddle' ~ PEC **ƛ̲wiłē* 'saddle'
- Bur. **=ltá-* > *tá-* /*=-ltá-* 'to put on (shoes, stockings)'; **holtáṣ* 'barefoot' ~ PEC **=ōmŁ̲V* 'to put on (trousers, shoes)' ~ Bsq **ortúć* 'barefoot; to take off (shoes, stockings)'

(c) Examples with Burushaski medial ***-lt-** *(~ -ld-)*:
- Bur. **baltí* 'front room, veranda' ~ PEC **bŭlƛ̲V* 'house'
- Bur. **díltar* 'buttermilk' ~ PNC **rħăƛwV̆* 'milk'
- Bur. **daltán-* 'to thresh' ~ PEC **=V̄rŁV* 'to thresh' ~ Bsq **laṙain* 'threshing floor'
- Bur. **=hált-* 'to wash' ~ PEC **=Vƛ̲Vn* 'to wash, pour; weep' (NCED 1023)
- Bur. **=ltVr-* 'to show' ~ PEC **ʔiŁ̲V* 'to look' (NCED 209)
- Bur. **malt-áṣ* 'butter' ~ PEC **nħĕƛ̲V* 'milk, dairy product'
- Bur. **=múltur* 'nostril' ~ PEC **wĕnƛ̣V* 'horn, head, nose' ~ Bsq **mutuṙ* 'snout, muzzle'
- Bur. **=yáldi-* 'part of the ribs', etc. ~ PEC **Hĕƛ̲V* ~ **HĕŁ̲V* 'armful'
- Bur. **(y)ált* 'spoon, scoop' ~ PEC **yă[l]ƛwV* 'wooden shovel' ~ Bsq *śa-*ɦarde* 'rake, fork'
- Bur. **yáltar* (H,N) 'upper leafy branches', etc. ~ PEC **ɦălƛ̲VłV* 'branch, pod' ~ Bsq **adaṙ* 'branch, knot', etc. < **ardaṙ*

(d) Examples with Burushaski final ***-l**:
- Bur. **tal* 'dove' ~ PEC **ƛ̲eƛ̣ē* 'a kind of bird'
- Bur. **=úl* 'belly, abdomen, bowels' ~ PEC **=ɨ̆raŁ̲V* 'stomach, rennet, abomasum'
- Bur. **=ýal-* 'to hear' ~ PNC **=eƛu* 'to hear' (NCED 411, CSG 46)

One East Caucasian language, Avar, partially converges with Burushaski in the development [t'] < PDC *ƛ̣.

- Avar *t'álu* 'rock, rocky plateau' (< PEC **ƛ̣ăłŭ*) ~ Bur **talí* 'slope (of a mountain)' (SCG 135)
- Avar (southern) *t'eb* 'millstone, whetstone' (< PEC **ƛ̣ĕṗV̄*) ~ Bur **tápi* 'stone terrace'
- Avar *t'eréna*-b 'thin' (< PNC **=iƛ̣ɨ̆lV*) ~ Bur. **tharén*-um 'narrow, tight'
- Avar *t'ínu* 'bottom' (< PNC **Hƛ̣ŏnŭ*) = Bur. **táno* 'rectum'
- Avar *t'ul* 'liver' (< PEC **Hƛ̣älV*) = Bur. **tal* 'stomach, belly'

Interesting as this may be, the Avar change differs from that of Bur in that it is restricted to PNC/PEC **ƛ̣(w)*, and in all positions (NCED 52), while the Bur change involves *all* lateral affricates, and only in *initial* position (see above). Some additional examples of Bur **t- / *-lt- / *-l* corresponding to PDC lateral affricates are given in Bengtson & Blažek (2011).

Conclusions

On the basis of about 100 lexical cognate sets, together with regular phonological patterning of the cognates, and shared irregular morphological patterns, a genetic relationship among Burushaski, the North Caucasian languages, Basque, and Yeniseian languages is the best hypothesis for explaining these similarities. On a competing hypothesis, that Burushaski has a close genetic relationship with Indo-European (Čašule 1998, 2003), see Bengtson (2000), Bengtson & Blažek (2011a, 2011b).

Cultural vocabulary shared by the same languages, including words for domestic sheep and goats, dairying, cultivated grain crops (and processes connected with them), and for other artifacts, suggests that the speakers of the proto-language ancestral to these languages (Euskaro-Caucasian or Macro-Caucasian) dispersed as early as 7000 to 9000 years BP in association with the spread of animal domestication and the cultivation of grain.[65]

Postscript (2014)

Since the original text of this paper was written, for the the Third Harvard Round Table in 2001, many new developments have taken place in Dene-Caucasian studies. The late Sergei A. Starostin, following up on the work of his colleague Ilia Peiros (1988) and myself, worked intensively during the last months of his life on integrating Burushaski into his Sino-Caucasian hypothesis (which up to that point had only included North Caucasian, Yeniseian, and Sino-Tibetan), greatly expanding the number of Burushaski-Dene-Caucasian etymologies, and more fully formulating the comparative phonology briefly sketched above. Before his untimely death in 2005 Starostin was able to integrate Burushaski into his *Sino-Caucasian Phonology* (2005a), *Sino-Caucasian Glossary* (2005b), and the Tower of Babel etymological databases (2005c, 2005d, 2007).

Since that time other members of the "Moscow School," especially George Starostin and Alexei Kassian, have further developed the Dene-Caucasian hypothesis; and credit must also be given to Václav Blažek and Merritt Ruhlen (see References). Using glottochronology, G. Starostin has proposed a new taxonomic model of Dene-Caucasian:

A. 'Sino-Dene' or 'Eastern Dene-Sino-Caucasian':
    A.1. Sino-Tibetan;
    A.2. Na-Dene;
B. 'Western Dene-Sino-Caucasian':
    B.1. Yeniseian + Burushaski;
    B.2. North Caucasian + Basque.

Node **A** ('Sino-Dene') essentially confirms the proposal made by Edward Sapir nearly a century ago. Node **B.1**, 'Burusho-Yeniseian', coincides with George van Driem's (2001) model, which he has called 'Karasuk', and has been further developed recently by Bengtson (2010). The approximate dates of these nodes are ~10,660 BCE (division of **A** from **B**), ~9000 BCE (division of **A.1** from **A.2**), slightly later (~8,330 BCE) for the split between **B.1** and **B.2**, and about two millennia later (~6,570 BCE) for the breakup of Burusho-Yeniseian.

---

[65] See **Postscript (2014)** for the glottochronological dates calculated by G. Starostin.

As to the Burusho-Yeniseian node, some shared semantic innovations were mentioned above:

- Bur. *_meṣ_ 'finger/toe' ~ PY *_bes_- 'finger' (vs. NC 'hoof, nail, claw')
- Bur. *_ć(h)argé_) 'flying squirrel' ~ PY *_saʔqa_ 'squirrel' (vs. NC 'weasel, mouse', Bsq 'mouse')
- Bur. *_khen_ 'flea' ~ PY *_qəʔń_ 'flea' (vs. NC 'louse, nit, worm', Bsq 'tick')

Several other close Burusho-Yeniseian isoglosses can be mentioned:

- Bur. *_=yek_ 'name' ~ PY *_ʔiɢ_ 'name'
- Bur. *_=reŋ_ 'hand' ~ PY *_ŕɔŋ_ 'hand'
- Bur. *_gan_ 'road' ~ PY *_kāŋ_- (~ *_gāŋ_-) '(hunting) path'
- Bur. *_phuiŋ_ 'nape (of neck); shoulder' ~ PY *_p[u]ym_- 'neck'
- Bur. *_toq_ 'mud' ~ PY *_təq_- 'mud, clay'

Among grammatical homologies one of the most obvious is the second person singular verbal affix, in both families inserted before the verb root:

> Bur. (2ps verbal affix) ***gu-/-ku-, gu-, go-/-ko-, goo-/-koo-*** ~ Ket (2ps verbal affix) ***k-, ku-***

Another grammatical homology is found in the demonstrative pronouns, with a common stem *_k_$^{(h)}$_i-_:

> Bur. (Y) *khin, khene* 'this' (hm-class) ~ Ket (Pak, Sur) *kīdə* 'this' (m.), etc.

Burushaski and Yeniseian have some similar interrogative pronouns made up of the elements *_bV_ + *_sV_:

> Burushaski: (Y) *besa, bese* 'why?', (H, N) *besan* 'what, which?', *bese* 'why?'
> Ket *biśέŋ / biśáŋ* 'where?', *biśśe* 'who?' (masc.), *bεśa* 'who?' (fem.)

The evidence for Burusho-Yeniseian is discussed in more detail in Bengtson (2010).

What about "Dené-Yeniseian"

Building on a proposal by Ruhlen (1998a) Edward Vajda (see References) has elaborated a "Dené-Yeniseian" hypothesis, attracting attention and acceptance from some 'mainstream' linguists, and in the popular press, e.g. Diamond (2011). At a recent Dene/Athabaskan Languages Conference (Berkeley, Calif., 2009) Vajda, George Starostin, and I presented reports about Dené-Yeniseian and Dene-(Sino-)Caucasian. Starostin and I asserted that the binary Dené-Yeniseian model is less adequate than the multilateral Dene-Caucasian one, and that Dené-Yeniseian is taxonomically inaccurate, since, in our view, Na-Dene is closer to Sino-Tibetan, and Yeniseian is closer to Burushaski, than Na-Dene and Yeniseian are to each other (see G. Starostin's diagram above). We maintained that the bigger, if somewhat messier, picture of a DC macrofamily is a better and more holistic approximation of the prehistoric reality. The relevant reports have since been published as Vajda (2010), Bengtson (2010) and G. Starostin (2010a). More recently, see further discussions by G. Starostin (2012) and Vajda (2012).

It is plain to see from the cultural lexicon discussed above that the Yeniseian languages share virtually none of the vocabulary pertaining to food production (domestic cattle, dairying, cultivated grain crops and their processing), with the other 'Western Dene-Caucasian' languages, Burushaski, Caucasian, and Basque. This is simply because the climate and landscape in the homeland of the Yeniseian peoples, in the Siberian taiga, are not conducive to food production, and presumably the Proto-Yeniseians either did not participate in the Neolithic revolution, or were forced by a new environment to abandon food production. The only exception (until Soviet rule) was the adoption of domestic reindeer by some Ket groups, from Samoyedic neighbors (Vajda 1998). So to some extent, at least, the apparent similarities between Yeniseian and Na-Dene in lexicon associated with sleds, snowshoes, birch-bark, and the like, are to be attributed not to taxonomic closeness but to similar environmental and cultural conditions.

Abbreviations

BN	bas-navarrais / behe nafarrera / Low Navarrese (Basque dialect)
Bzk	Bizkaian = Biscayan (Basque dialect)
Bzt	Baztanese (Basque dialect)
c	common Basque
DC	Dene-Caucasian
EC	East Caucasian (= Northeast Caucasian or Nakh-Dagestanian)
H	Hunza (Burushaski dialect)
Lap	Lapurdian = Labourdin (Basque dialect)
N	Nager (Burushaski dialect)
PA	Proto-Athabaskan
PAE	Proto-Athabaskan-Eyak
PDC	Proto-Dene-Caucasian
PEC	Proto-East Caucasian
PNC	Proto-(North) Caucasian
PWC	Proto-West Caucasian
PY	Proto-Yeniseian
Rnc	Roncalese (Basque dialect)
WC	West Caucasian (= Northwest Caucasian or Abkhaz-Adygean)
Y	Yasin (Burushaski dialect) = "Werchikwar"
Zub	Zuberoan = Souletin (Basque dialect)

Buck	Buck (1949)
CDIAL	Turner (1962-66)
LACUS	Linguistic Association of Canada and the United States
NCED	Nikolayev & Starostin (1994)
SCG	Sino-Caucasian Glossary (Starostin 2005b)
SCP	Sino-Caucasian Phonology (Starostin 2005a)

Sino-Dene	Bengtson (1994)
SKJa	*Slovar' kavkazskix jazykov* (Klimov & Xalilov 2003)
TOB	Tower of Babel Etymological Databases http://starling.rinet.ru/cgi-bin/main.cgi?flags=eygtnnl

References

Anthony, David W. & Dorcas R. Brown. 2011. "The Secondary Products Revolution, Horse-Riding, and Mounted Warfare." *Journal of World Prehistory* 24: 131–160.

Bengtson, John D. 1994. "Edward Sapir and the 'Sino-Dene" Hypothesis." *Anthropological Science* 102: 207-230. http://www.nostratic.ru/books/(219)bengtson%20-%20sapir.pdf.

Bengtson, John D. 1997a. "Ein Vergleich von Buruschaski und Nordkaukasisch." *Georgica* 20.88-94.

Bengtson, John D. 1997b. "Basque and the Other Dene-Caucasic Languages." In Melby (1997), pp. 137-148.

Bengtson, John D. 2000. Review of Čašule 1998. *History of Language* 6.1: 22–26.

Bengtson, John D. 2001. "Genetic and Cultural Linguistic Links between Burushaski and the Caucasian Languages and Basque." Paper given at the 3rd Harvard Round Table on the Ethnogenesis of South and Central Asia. Harvard University, May 2001. http://www.nostratic.ru/books/%28299%29bengtson%20-%20BurshRoundtable.pdf.

Bengtson, John D. 2002. "The Dene-Caucasian Noun Prefix **s-*." *The Linguist's Linguist: A Collection of Papers in Honour of Alexis Manaster Ramer,* ed. by F. Cavoto, vol. 1, 53-57. Munich: LINCOM EUROPA.

Bengtson, John D. 2008a. "Materials for a Comparative Grammar of the Dene-Caucasian (Sino-Caucasian) Languages." *Aspects of Comparative Linguistics*, vol. 3, 45-118. Moscow: RSUH Publishers. http://starling.rinet.ru/Texts/dene_gr.pdf.

Bengtson, John D. (ed.) 2008b. *In Hot Pursuit of Language in Prehistory: Essays in the four fields of anthropology. In honor of Harold Crane Fleming*. Amsterdam/Philadelphia: John Benjamins.

Bengtson, John D. 2008c. "The languages of northern Eurasia: Inference to the best explanation." In Bengtson 2008b, 241-262.

Bengtson, John D. 2010. "'Dene-Yeniseian' and the Rest of Dene-Caucasian. Part 3: The Burusho-Yeniseian (Karasuk) Hypothesis. Part 4: Burusho-Dene." *Working Papers in Athabaskan Languages 2009. Alaska Native Language Center Working Papers*, No. 8, 1-18.

Bengtson, John D., & Václav Blažek. 2011a. On the Burushaski-Indo-European hypothesis by I. Čašule. *Journal of Language Relationship* 6: 25-63.

Bengtson, John D., & Václav Blažek. 2011b. Comments on Ilija Čašule's "Correlation of the Burushaski Pronominal System with Indo-European and Phonological and Grammatical Evidence for a Genetic Relationship." *Journal of Indo-European Studies* 40.1/2: 156-161.

Berger, Hermann. 1956. "Mittelmeerische Kulturpflanzennamen aus dem Burušaski." *Münchener Studien zur Sprachwissenschaft* 9.4-33.

Berger, Hermann. 1959. "Burušaski Lehnwörter in der Zigeunersprache." *Indo-Iranian Journal* 3:1.17-43.

Berger, Hermann. 1974. *Das Yasin-Burushaski (Werchikwar).* Wiesbaden: Harrassowitz.

Berger, Hermann. 1998. *Die Burushaski-Sprache von Hunza und Nager.* 3 vols. Wiesbaden: Harrassowitz.

Blažek, Václav. 1995. "Towards the Position of Basque: A reply to Trask's critique of the Dene-Caucasian hypothesis." *Mother Tongue* 1: 104-110.

Blažek, Václav. 1999. Some remarks on 'A Comparison of Basque and (North) Caucasian Basic Vocabulary" by John D. Bengtson." *Mother Tongue* 5: 59-61.

Blažek, Václav, & John D. Bengtson. 1995. "Lexica Dene-Caucasica." *Central Asiatic Journal* 39.1: 11-50, 39.2: 161-164. http://www.nostratic.ru/books/(223)DeneCaucasicaCAJ.pdf.

Bleichsteiner, R. 1930. "Die werschikisch-buruschkische Sprache im Pamir-Gebiet und ihre Stellung zu den Japhetitensprachen des Kaukasus." *Wiener Beiträge zur Kunde des Morgenlandes* 1.289-331.

Bouda, Karl. 1948. "Baskisch und Kaukasisch." *Zeitschrift für Phonetik* 2.182-202, 336-352.

Bouda, Karl. 1949. *Baskisch-Kaukasische Etymologien.* Heidelberg.

Bouda, Karl. 1950. "Die Sprache der Buruscho." *Eusko-Jakintza* 4.37-50, 337-346.

Bouda, Karl. 1951. "Ein jüngst ermittelte archaische Sprachgruppe in Asien und Europa." *Germanisch-Romanische Monatsschrift* 32.129-142.

Bouda, Karl. 1954. "Burushaski Etymologien." *Orbis* 3.2: 228-230.

Bouda, Karl. 1964. "Burushaski Etymologien II." *Orbis* 13.2: 604-609.

Buck, Carl Darling. 1949. *A Dictionary of Selected Synonyms in the Principal Indo-European Languages.* Chicago: University of Chicago Press.

Čašule, Ilija. 1998. *Basic Burushaski Etymologies: The Indo-European and Paleo-Balkanic Affinities of Burushaski.* LINCOM Etymological Studies 01. Munich: LINCOM Europa.

Čašule, Ilija. 2003. "Evidence for the Indo-European Laryngeals in Burushaski and Its Genetic Affiliation with Indo-European." *Journal of Indo-European Studies* 31.1/2: 21–86.

Chirikba [Čirikba], Vyacheslav. 1996. *Common West Caucasian.* Leiden: Research School CNWS.

Colledge, S., J. Conolly, K. Dobney, K. Manning, S. Shennan (eds.). 2013. *The Origins and Spread of Domestic Animals in Southwest Asia and Europe.* Walnut Creek, Calif.: Left Coast Press.

Diamond, Jared. 2011. "Linguistics: Deep relationships between languages." *Nature* 476 (7360): 291-292.

Driem, George van. 2001. *Languages of the Himalayas: An Ethnolinguistic Handbook of the Greater Himalayan Region, containing an Introduction to the Symbiotic Theory of Language.* 2 vols. Leiden: Brill.

Gerbault, P., R. Bollongino, J. Burger, M.G. Thomas. 2013. "Inferring Processes of Neolithic Gene-Culture Co-Evolution Using Genetic and Archaeological Data: The Case of Lactase Persistence and Dairying." In Colledge, et al. (eds.) 2013: 37-48.

Kassian, Alexei. 2010a. "Hattic as a Sino-Caucasian language." *Ugarit-Forschungen.* Bd. 41. 2009, pp. 309-448.

Kassian, Alexei. 2010b. Review of A. Fournet, A.R. Bomhard, *The Indo-European Elements in Hurrian* (2010). *Journal of Language Relationship* 4.199-206.

Kassian, Alexei. 2010c. "Hurro-Urartian from the lexicostatistical viewpoint." *Ugarit-Forschungen.* Bd. 42. 383-451. 2010.

Klimov, G.A., & M.Š. Xalilov. 2003. *Slovar' kavkazskix jazykov: Sopostavlenie osnovnoj leksiki.* [*Dictionary of Caucasian Languages: comparison of the principal lexis.*] Moscow: Vostočnaja literatura.

Leer, Jeff. 1993. "AET Comparative Lexical Database." Ms. ANLC.

Marshall, F. & L. Weissbrod. 2011. "Domestication Processes and Morphological Change through the Lens of the Donkey and African Pastoralism." *Current Anthropology* 52 (S4): S397-S413, October 2011.

Melby, Alan K. (Ed.) 1997. *The Twenty-Third LACUS Forum.* Chapel Hill, N.C.: LACUS.

Outram, A.K., N.A. Stear, R. Bendrey, S. Olsen, A. Kasparov, V. Zaibert, N. Thorpe, R.P. Evershed. 2009. "The Earliest Horse Harnessing and Milking." *Science* 323.5919: 1332-1335.

Peiros, Ilia. 1988. "Sinokavkazskaja teorija i burušaskij jazyk." *Problemy izučenija sravnitel"no-istoričeskogo jazykoznanija i lingvističeskoj istorii o vostoke i juge Azii*, 214-227. Moscow: Institute of Far Eastern Studies, Academy of Sciences of the U.S.S.R.

Ruhlen, Merritt. 1995. "Comments on R. L. Trask's Critique: Is Basque an Isolate?" *Mother Tongue* 1: 149-156.

Ruhlen, Merritt. 1996a. "Letter from Merritt Ruhlen to Larry Trask." [Response to Trask 1996c.] *Mother Tongue* 2: 137-139.

Ruhlen, Merritt. 1996b. "Une nouvelle famille de langues: le déné-caucasien." *Pour la Science* (Dossier, Octobre): 68–73.

Ruhlen, Merritt. 1998a. "The Origin of the Na-Dene." *Proceedings of the National Academy of Sciences* 95: 13994-13996. http://www.merrittruhlen.com/files/1998c.pdf.

Ruhlen, Merritt. 1998b. "Dene-Caucasian: A New Linguistic Family." K. Omoto and P. V. Tobias (eds.), *The Origins and Past of Modern Humans – Towards Reconciliation*, 231-246. Singapore: World Scientific.

Ruhlen, Merritt. 2001. "Il Dene-caucasico: una nuova famiglia linguistica." *Pluriverso* 2: 76–85.

Starostin, George. 2010a. "Dene-Yeniseian and Dene-Caucasian: Pronouns and Other Thoughts." *Working Papers in Athabaskan Languages 2009. Alaska Native Language Center Working Papers*, No. 8, 107-117.

Starostin, George. 2010b. "Preliminary lexicostatistics as a basis for language classification: a new approach." *Journal of Language Relationship* 3: 79-117.

Starostin, George. 2012. "Dene-Yeniseian: a critical assessment." *Journal of Language Relationship* 8: 117-138.

Starostin, Sergei A. 2005a. *Sino-Caucasian [Phonology]*. Ms. http://starling.rinet.ru/Texts/scc.pdf.

Starostin, Sergei. 2005b. *Sino-Caucasian [Glossary]*. Ms. http://starling.rinet.ru/Texts/glossary.-pdf.

Starostin, Sergei. 2005c. North Caucasian Etymology (Database). Tower of Babel. http://starling.rinet.ru/cgi-bin/main.cgi?root=config

Starostin, Sergei. 2005d. Burushaski Etymology (Database). Tower of Babel. http://starling.rinet.ru/cgi-bin/main.cgi?root=config

Starostin, Sergei. 2007. Sino-Caucasian Etymology (Database). Tower of Babel. http://starling.rinet.ru/cgi-bin/main.cgi?root=config

Toporov, Vladimir. 1971. "Burushaski and Yeniseian Languages: Some Parallels." *Travaux linguistiques de Prague* 4: 108-125.

Turner, R. L. 1962-66. *A comparative dictionary of Indo-Aryan languages*. London: Oxford University Press.

Vajda, Edward. 1998. "The Kets and Their Language." *Mother Tongue* 4: 4-16.

Vajda, Edward. 2000a. "Yeniseian and Na-Dene: Evidence for a genetic relationship." Paper presented at the 38th Conference on American Indian Languages (SSILA), Chicago, January 2000.

Vajda, Edward. 2000b. "Yeniseian and Athabaskan-Eyak-Tlingit." Paper presented at the Linguistics Department Colloquium, University of British Columbia, March, 2000.

Vajda, Edward. 2000c. "Ket verb morphology and its parallels with Athabaskan-Eyak-Tlingit: evidence of a genetic link." Paper presented at the Athabaskan Language Conference, Moricetown, British Columbia, June 9, 2000.

Vajda, Edward. 2005. "Yeniseic and Na-Dene." Talk presented at Max Planck Institute for Evolutionary Anthropolo-gy, Leipzig, Germany, Dec. 8, 2005.

Vajda, Edward. 2008. "The Siberian origins of Na-Dene languages." Paper presented at the Dene-Yeniseic Work Session, Fairbanks, Feb. 26-27, 2008, and at the Annual Meeting of the Alaskan Anthropological Society, Anchorage, Feb. 28, 2008.

Vajda, Edward. 2010. "A Siberian link with Na-Dene languages." *The Dene-Yeniseian Connection. Anthropological Papers of the University of Alaska*, 5.1-2: 33-99.

Vajda, Edward. 2012. "The Dene-Yeniseian connection: a reply to G. Starostin." *Journal of Language Relationship* 8: 138-152.

Witzel, Michael. 1999b. "Early Sources for South Asian Substrate Languages." *Mother Tongue* Special Issue: October 1999: 1-71.

Book Reviews

E.J. Michael Witzel: *The Origins of the World's Mythologies.*
Oxford - New York: Oxford University Press. 2011.

ISBN 978-0-19-536746-1 (hardcover) / ISBN 978-0-19-981285-1 (paperback), pp. xxii + 665.

Reviewed by Václav Blažek
Department of Linguistics and Baltic Studies
Faculty of Arts, Masaryk University, Brno
Czech Republic

This long-expected monograph by Michael Witzel, Wales professor of Sanskrit at Harvard University, is unique for several reasons. The author brings forth a revolutionary stratigraphy of myths, determined by two chronological and geographical layers, an earlier southern, called Gondwanan, and a later northern, called Laurasian. Both terms were chosen by the author on the analogy of geological designations of the supercontinents Gondwana and Laurasia, separated $200\text{-}180 \cdot 10^6$ years ago in the mid-Mezozoic era. Naturally, the author does not try to date the oldest myths to this period; his chronological limit is *c.* $160 \cdot 10^3$ years (p. 372), when modern humans appeared in Africa.

M. Witzel summarizes the typical features of the Laurasian mythological traditions as follows (p. 64):

1. Primordial waters / chaos / 'nonbeing'
2. Primordial egg / giant.
3. Primordial hill or island.
4. (Father) Heaven / (Mother) Earth and their children (4 or 5 generations / ages).
5. Heaven is pushed up (and origin of Milky Way).
6. The hidden sun light revealed.
7. Current gods defeat or kill their predecessors.
8. Killing the 'dragon' (and use of heavenly drink), fertilization of the earth.
9. Sun deity is the father of humans (or just of 'chieftains').
10. First humans and first evil deeds (often, still by a demi-god), origin of death / the flood.
11. Heroes and nymphs.
12. Bringing of culture: fire / food / culture by a culture hero or shaman; rituals.
13. Spread of humans / emergence of local nobility / local history begins.
14. Final destruction of humans, the world (and) the gods (variant of the Four Ages theme).
15. (A new heaven and a new earth).

The mythological traditions defined as Gondwanan are not as rich, although in several features they agree with the Laurasian ones (pp. 322-323):

1. In the beginning : heaven, earth (and the sea) already exist.
2. A High God lives in heaven, or on earth, or ascends to heaven later. This highest being has been described as the Rainbow Snake.
3. Series of lower gods, often children of High God, act as tricksters and culture heroes.
4. Primordial period ended by some evil deed of son of High God (or by humans).

5. Humans are created from trees and clay (or rock); occasionally, descend directly from the gods / totem ancestors.
6. Humans act haughtily or make a mistake; punishment by a great flood; humans reemerge in various ways. (An end to the world is missing.)

Some concepts seem to be universal, and M. Witzel designates them as Pan-Gaean:

1. High god.
2. Creation of heaven and earth.
3. Creation of humans from clay or from a tree.
4. Cultural hero or trickster.

The following comments concentrate on the linguistic implications of the book under review.

There are fascinating linguistic traces of the Rainbow Snake (cf. the map by Berezkin 2013, 115) and related supreme deities along the coast of the Indian ocean. Witzel (pp. 310, 326, 351, 352) mentions the Andamanese creator deity *Puluga* or *Biliku*, often identified with the northeast monsoon. The concrete forms and their sources are as follows:

> Great Andaman: Bea *púlúga-da*, Bale *púlúga*, Puchikwar *bilik-da*, Juwoi *bílak-lek̲íle*, Kol *bílak-che*, Bojigiab *bílak-da*, Kede *bílke*, Chariar *bílek ke* 'god' (Portman 1898, 68-69; 1887, 34-35), Jeru *biliku* 'a mythical being' (Brown 1914, 40, 51);
> Little Andaman: Önge *Öluga* 'mythical being identified with the northeast monsoon, name of the Monitor lizard' (Brown 1914, 40, 51) = *ölugé* 'to thunder' (Portman 1887, 83), *ulugé* 'god' (Portman 1887, 34-35).

This common Andamanese god was apparently specialized in thundering, to judge from the following designations for 'thunder':

> Great Andaman: Bea *púlúga-lá gāū̆rawa-ké*, Bale *púlúga-lé kúrúdá-ké*, Puchikwar *bílak-le gāū̆rawa-ke*, Juwoi *bílak-le ťrémè-che*, Kol *bílak-ke pàrak-le* 'thunder' (Portman 1898, 166-67), versus Bea *górawá ké* 'to thunder' (Portman 1887, 83).

The closest external cognate may be identified in Australia in one non-Pama-Nyungan family from the eastern Arnhem Land: proto-Gunwinyguan **polong* ~ **polung* 'rainbow serpent' > Dalabon, Jawoyn, Rembarrnga *polung*, Mangarrayi *polokpan* (Harvey 2003, 255, #893). Other possible Australian cognates were collected by Riccardo Gatti more than a century ago:

> non-Pama-Nyungan: Tangkic: Jakula *pargi-gi* 'god' (Gatti 1906, 6: Jakula + Andamanese) || Pama-Nyungan: Lower Murray: Ngintait *poorache*; Bagandji: Maraura *poorook*; Maric: Wadjabangai *baringa*; Waka-Kabic: Bayali *baroongi*; Central East Coast: Bigumbil *booringa*, *booronga*, Yugumbal *boorongi* 'thunder' (Gatti 1906, 31: Pama-Nyungan + Andamanese). Related are perhaps also Karadjerri (Pama-Nyungan family Marrngu on the northwest coast) *Bulaing* 'a tribal creator-goddess' and Warramunga / Warumungu (Pama-Nyungan isolate in south from the Gulf of Carpentaria) *wollunqua* 'rainbow serpent' (Mills 1918, 211).

Gatti (1906, 31) also added Tasmanian (glossary by Peron) *bura* 'thunder'.

In his introduction to Gatti 1906, Alfredo Trombetti (pp. ix-x) also added hypothetical parallels from Sub-Saharan Africa:

> (i) Oromo of Wellega *bulguu* 'ogre, cannibal (with four eyes)' (Gragg 1982, 67) = *bulgu* 'cannibal or man-eater believed to have four eyes' (Tilahun Gamta 1989, 95) =

bulgu 'antropofago, cannibale; creazione della fantasia e della superstizione locale' (Borello 1995, 66) = Boraana *bulguu* 'cannibal, bogey man' (Stroomer 1995, 159).

This term stands isolated within Cushitic (and Afroasiatic in general), and cannot be explained as an Ethio-Semitic loan, so its origin in some pre-Cushitic substratum seems quite probable.

(ii) Bantu supreme deity, documented in the zones EFGMNP[1], especially in Eastern Africa: Nyamwezi *mu-lungu* 'tribal creator- & sky-god watching over the earth', Kikuyu *mu-rungu* 'supreme god', Chiuta *mu-lengi* 'tribal creator', Kinga pl. *emị-luṅgu* 'forefathers of the common people', Swahili *mu-uṅgu*, pl. *mi-uṅgu* 'god', Herero *ka-ruṅga* 'O god!' Meinhof & Warmelo (1932, 212) reconstructed p-Bantu **-luṅgu*. Guthrie (1971) & his followers reconstruct p-Bantu **dòngò*[2] 'god'.

The missing initial labial syllable, in comparison with the preceding Cushitic examples, may be ascribed to reinterpretation in the frame of the noun-class system. Rather peculiar is the semantic determination of the class prefix pair sg. *mu-* vs. pl. *mi-*; in Swahili, for example, where both sg. & pl. operate, this pair is reserved for plants (e.g., *m(u)-ti* 'tree' / *mi-ti* 'trees').

Later Trombetti (1918-19, 35/363) abandoned the Bantu examples and added the following comparanda from the Niger-Congo languages of the Gulf of Guinea: Lefana (= Lelemi in Togo & Ghana; Kwa family) *Burukú* 'name of a deity' and Sobo (Edoid; south Nigeria) *sono-blūgwé* 'god'. It is possible to add Fon (Gbe group of the Volta-Niger family; Burkina Faso) *Nana Buluku*[3] 'androgynous supreme Creator of the Universe and all that exists in it', and Yoruba *Nana Burukú*[4] 'great grandmother, connected with rivers and oceans; she is the ultimate mother of the waters, especially the sweet river waters'

It is tempting to speculate about any relation to the deity celebrated by Ngbandi (Adamawa-Ubangi family of the Niger-Congo macro-family), called *Mbongo* 'a river-god and one of the seven deities invoked at sun rise every morning; the creator god of all tribal people who is said to reside in black waters' (Jordan 1993, 161).

The tree-diagram shown below [**Figure 1**] depicts hypothetical genealogical relations among the world's linguistic macro-phyla, with a preliminary chronological scale. The double line cutting across branch lines is a mythological "Wallace Line," separating the predominant Laurasian motifs from more archaic Gondwanan ones.

The 'Gondwanan' motifs are typical of people speaking Khoisan and Congo-Saharan languages in Sub-Saharan Africa, and Indo-Pacific and Australian languages in Australasia, probably descendants of the first out-of-Africa migration. But they appear more frequently than rarely also in two populations speaking Dravidian and Austronesian languages, whose proto-languages are affiliated with the Nostratic and Austric macro-families respectively, and both are connected with the later out-of-Africa migrations (see **Figure 1**). This contradiction is explainable on the basis of a strong influence of the pre-Dravidian and pre-Austronesian / pre-Austric substrata. The following examples may serve as illustration[5]:

---

[1] http://goto.glocalnet.net/mahopapers/nuglonline.pdf

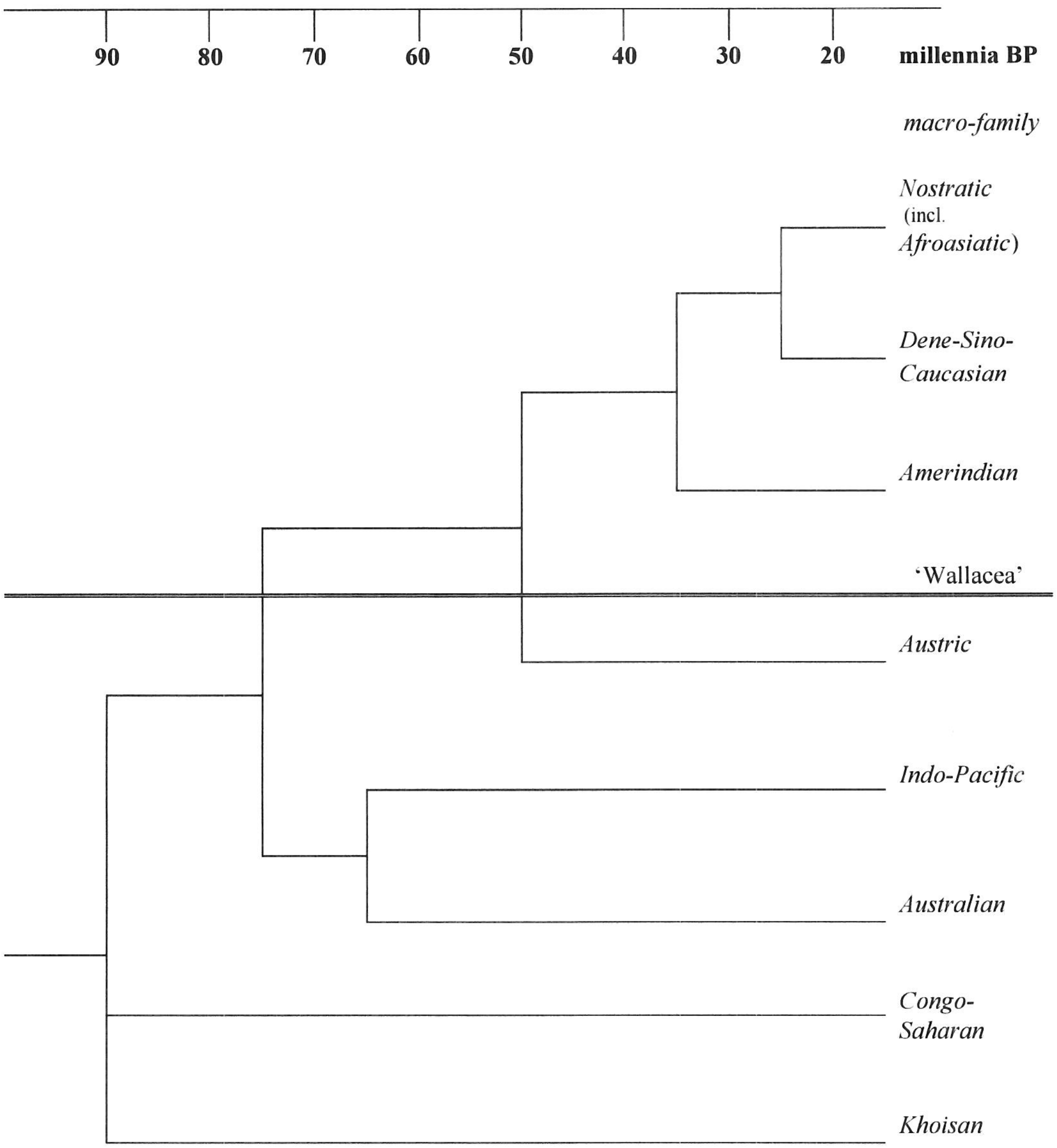
[2] http://www.africamuseum.be/collections/browsecollections/humansciences/blr/results_main

[3] http://en.wikipedia.org/wiki/Nana_Buluku

[4] http://myoruba.tumblr.com/post/78006900721/nana-buruku-is-the-great-grandmother-to-all-of-the

[5] Bengtson's (1994) stimulating comparisons between Austric, Indo-Pacific and Australian can be reinterpreted in this areal sense. Before the settlement of Austronesians in the Indonesian & Philippine archipelagoes, and other Austric populations on the coast of Southeast Asia, these territories were not uninhabited. The people living there probably spoke languages related to Indo-Pacific, whose remnants

Figure 1: Hypothetical genealogical relations among the world's linguistic macro-phyla, with a preliminary chronological scale. The double line [==] is a mythological "Wallace Line," separating the predominant Laurasian motifs from more archaic Gondwanan ones.


90
80
70
60
50
40
30
20
millennia BP
macro-family
Nostratic (incl. Afroasiatic)
Dene-Sino-Caucasian
Amerindian
'Wallacea'
Austric
Indo-Pacific
Australian
Congo-Saharan
Khoisan

STAR: (Trombetti 1918-19, 51/379)

- Common Australian **biṇḍiri* 'star' (Capell 1979, 487, 584) > Pama-Nyungan: South-West: Gascoyne R.-Pilbara: p-Ngayarda (northwest coast of Australia) **piṇṭi.ri* 'star' > Ngaluma, Jindjibandi *piṇṭiri* id. (= 10 *pindeeree* by Curr I), besides Panjima *piṭa*, North Yinggarda (Mantharda subgroup) *puṭa* id. (O'Grady 1966, 101, #176; 115, #845) | Moore R.-Gascoyne R.: Parti-Maya: 27 Widi *buʳndo*, 28 Muliara *bondar* | Spencer Gulf Basin: 65 Nukuna *bundi* || Greater Maric (Queensland): 137 Iningai *boodtha*, 144 Barna, 152 Wadjabangai, 153 Kuungkari *bootoo*, 158 Karingbal *boodoo & boothoo* 'star', 150 Kangulu *boodthou*, 151 Jambina *buttho*, 155 Tambo *boodtha*, 156 Kairi *boodthoe*, 127 Bindal *botho*, 130+131 Ilba *buttu & buthi* ||

---

in Timor and North Halmahera exist till the present, and Australian. An analogous situation was probably valid for the Indian peninsula (cf. Blažek 2006).

Kalkatungic: 101 Jalanga _booderoo_ (Australian forms with numbers are quoted after Curr; Gatti 1906, 29: Maric + Andamanese; Trombetti 1918-19, 51: Pama-Nyungan + Austronesian + North Dravidian).

- Tasmanian _piterina_ 'sun' & _potena_ 'star' (glossary of Lhotsky; Gatti 1906, 27 compared _potena_ with Widugari {Pilbara-Ngayarda subgroup of the Gascoyne R.-Pilbara R. group of South-West Pama-Nyungan} _peta_ 'star').
- Andamanese: Great Andaman: Bea _bódó-da_, Bojigiab _pute_ 'sun' (Portman 1887, 78), Bale _bāūdo_, Puchikwar _púte-da_, Juwoi _púte-leḵíle_, Kol _pútè-che_ id. (Portman 1898, 160-61).
- Austronesian *_bi(n)tuhən_ (D 1929) = *_bituqen_ (Dyen 1953a) 'star'.
- Dravidian: North: Malto _binḍke_, Kurukh _bīnkō_ 'star' (DEDR #4876: *_mīṉ_, but cf. Tamil _viṇ_, Telugu _vinu_ 'sky', Malto _biṇye_ 'god of thunder and lightning' DEDR #5396).

WOMAN: (Bengtson 1994, 74)

- Australian: non-Pama-Nyungan: Daly R. (N): Mullukmulluk _pen_ 'vulva' (Tryon 1974, 269) ||| proto-Pama-Nyungan *_panyji.l_ 'woman, female' > South-West: Ngayarda: Yindjibarndi _pinhthi_ 'woman's' | Paman or Yidinic: Yabugay _panyji.l_ 'woman' (O'Grady 1990, 240) | Maric: Guwa _buña(na)_ 'woman' | Kulinic: Bundyil _buno_ | Karnic: Karuwali _punja_, _punga_ 'woman' (Schmidt).
- Indo-Pacific (Greenberg 1971, 858): Trans-New Guinea: Banak _pana_, _fana_ 'woman, wife'; Dem _pani_ 'female'; West Papuan: Aitinjo _finya_ 'women'; Mogetemin _fanya_ 'woman', etc.
- Austronesian *_binay_ (Dempwolff) ~ *_binəy_ (Dyen).
- Dravidian *_peṇ_ 'woman' (DEDR 4395: I, II, III, IV, V, VI).

STONE: (Blažek 2006, #68)

- Australian: non-Pama-Nyungan: Gunwinyguan *_kal-_ 'mountain, hill' (Peiros 428, 441) ||| Pama-Nyungan: Mabuiag _kula_; Mbambylmu ('Princesse Charlotte Bay') _kūla_; Gumbaynggiric _kullam_, etc. 'stone' (Schmidt).
- Austronesian *_kaḷaṅ_ 'rock' (Dempwolff 1929).
- Dravidian *_kal_ 'stone' (DEDR 1298: I, II, III, IV, V, VI, VIII).

WIND: (Blažek 2006, #62)

- Australian: non-Pama-Nyungan: Daly R.: Marityaben, Marengar _wirir_ 'wind' (Tryon 1974, 105); Pama-Nyungan: South-West: p-Ngayarda *_waḷpa_; Western Desert: Warnman & Yulparitja _waḷpa_ id. (O'Grady 1966, 116) | Aranda _olupa_ | Yalanjic: Gugu Yimidhir _walbun_ | NWKulin _willa_ | Kuri: Yugambal _wollar_, etc. 'wind' (Schmidt).
- Andamanese: Bea _wúl-nga-da_ 'wind' (Portmann 1898, 186).
- Dravidian *_vaḷi_ 'wind' (DEDR 5312: I, III, V, VI).
- Austronesian *_baliŋ_ 'wind' (Dempwolff).

Summing up, _The Origins of the World's Mythologies_ is not only a unique encyclopedia of comparative mythology, genetics and anthropology, but also some very inspirational reading, likely to stimulate new research in the field of comparative linguistics.

References

Bengtson, John D. 1994. Macro-Australic. *Mother Tongue Newsletter* 23, 73-75.

Berezkin, Jurij E. 2013. *Afrika, migracii, mifologija. Areal'nye rasprostranenija fol'klornyx motivov v istoričeskoj perspektive*. Sankt-Peterburg: Nauka.

Blažek, Václav. 2006. Was there an Australian Substratum in Dravidian? *Mother Tongue* 11, 2006, 275-294.

Borello, P. Mario. 1995. *Dizionario Oromo-Italiano*, ed. by Hans-Jürgen Sasse & Paolo Tablino. Hamburg: Buske.

Brown, A.R. 1914. Notes on the Languages of the Andaman Isleands. *Anthropos* 9, 36-52.

Capell, Arthur. 1979. *The History of Australian Languages: A First Approach*. In: *Australian Linguistic Studies*, ed. Stephen A. Wurm. Canberra: Australian National University (Pacific Linguistic Series, C-No. 54), 419-619.

Curr, Edward M. 1886-1887. *The Australian race : its origin, languages, customs, place of landing in Australia and the routes by which it spread itself over the continent*, I-IV. Melbourne: Ferres - London: Trübner.
II: https://archive.org/details/cu31924026093827
III: http://www.archive.org/stream/cu31924026093835#page/n0/mode/2up

DEDR = *A Dravidian Etymological Dictionary*, ed. by Thomas Burrow & Murray B. Emeneau. Oxford: University Press 1984.

Gatti, Riccardo. 1906-1908. *Studi sul gruppo linguistico Andamanese-Papua-Australiano*, I, II Bologna: Beltrami.

Gragg, Gene B. 1982. *Oromo Dictionary*. East Lansing: Michigan State University - African Studies Center.

Greenberg, J.H. 1971. The Indo-Pacific hypothesis. In: *Current Trends in Linguistics*, Vol. 8: *Linguistics in Oceania*. The Hague - Paris: Mouton, 807-871.

Guthrie, Malcolm. 1971. *Comparative Bantu*. Farnborough: Gregg International.

Harvey, Mark. 2003. An initial reconstruction of Proto-Gunwinyguan phonology. In: *The non-Pama-Nyungan languages of northern Australia: comparative studies of the continent's most linguistically complex region*, ed. by Nicholas Evans. Canberra: The Australian National University (Pacific Linguistics 552), 205-268.

Jordan, Michael. 1993. *Encyclopedia of Gods*. New York: Facts On File.

Meinhof, Carl. & van Warmelo, N.J. 1932. *Introduction to the Phonology of the Bantu Languages*. Berlin: Reimer - Vohsen.

Mills, Laos. 1918. *Prehistoric Religion*. Washington: Capital publishers.

O'Grady, Geoffrey N. 1966. Proto-Ngayarda phonology. *Oceanic Linguistics* 5, 71-130.

O'Grady, Geoffrey N. 1990. Pama-Nyungan: The tip of the lexical iceberg. In: *Studies in Comparative Pama-Nyungan*, ed. by G.N. O'Grady & D.T. Tryon. Canberra: Pacific Linguistics, C-111, 209-259.

Portman, M.V. 1887. *A Manual of the Andamanese Languages*. London: Allen.

Portman, M.V. 1898. *Notes on the languages of the South Andaman Group of Tribes*. Calcutta: Government Printing.

Schmidt, Wilhelm. 1912-18. Die Gliederung der Australischen Sprachen. *Anthropos* 7, 8, 9, 12-13.

Stroomer, Harry. 1995. *A Grammar of Boraana Oromo (Kenya)*. Köln: Köppe.

Tilahum Gamta. 1989. *Oromo-English Dictionary*.Addis Ababa: AAU Printing Press.

Trombetti, Alfredo. 1914-19. Saggi di glottologia generale comparata: III. Comparazioni lessicali. *Memorie della R. Accademia della Scienze dell' Istituto di Bologna*, Ser. I, T. IX-X; Ser. II, T. I-IV.

Tryon, D.T. 1974. *Daly Family Languages, Australia*. Canberra: Pacific Linguistics Series C - No. 32.

Electronic publications

Bantu classification: http://goto.glocalnet.net/mahopapers/nuglonline.pdf

Acknowledgement: The present study was prepared thanks to a grant from the The Czech Science Foundation (GAČR), P406/12/0655.

Václav Blažek <blazek@phil.muni.cz>
Department of Linguistics and Baltic Studies
Faculty of Arts, Masaryk University, Brno, Czech Republic

Michael Fortescue: *Language Relations across Bering Strait: Reappraising the Archaeological and Linguistic Evidence.*

Open Linguistics Series. London and New York: Cassell, 1998.
Pp. x + 307.

Reviewed by Peter A. Michalove[6]
University of Illinois

This book explores the relationships among four language families, Uralic, Yukagir, Chukchi-Kamchatkan, and Eskimo-Aleut (collectively referred to here as the Uralo-Siberian or US languages). While Fortescue is probably most widely known for his work in Eskimo-Aleut, he has worked with all of these languages to varying degrees, and so brings a deep knowledge of the material to this study.

Fortescue appears to be ambivalent about the nature of this relationship, referring to a "mesh, if not an actual genetic stock" (p. 3 and passim), explaining later (p. 28) that this term refers to

> any degree of historical relatedness between a group of (once) geographically related languages linked by lexical and/or phonological or structural 'family resemblance'. The possibilities range from *Sprachbunds* which have drawn unrelated languages into their orbit, through interlocking chains of languages where the ends are unrelated but where there is considerable overlap and even language mixing, to traditional language meshes that exclusively involve languages of a single family. Another, crucial intermediate possibility is the situation where all the ingredient languages are ultimate derived from a single ancestral protolanguage but the time depth is too great to prove it.

Yet elsewhere Fortescue speaks routinely in the terminology of demonstrable genetic language relations. For example he refers (p. 77) to, "elements inherited from Proto-US." Similarly, when he speaks of the undeniable presence of loans among these languages, he sees the loans, "[a]s opposed to genetically common lexical material apparently shared by the various Uralo-Siberian languages."

The introductory chapter states Fortescue's thesis, that there is a demonstrable historical relationship among Uralic, Yukagir, Chukchi-Kamchatkan, and Eskimo-Aleut. Fortescue's goal is to pin down as much as possible of the historical connections among these languages by supplementing the results of traditional comparative linguistics with typological and archaeological findings.

Chapter Two discusses previous attempts to relate various combinations of these languages, as well as the ongoing debate over Chukchi-Kamchatkan itself. Most scholars today accept a Chukchi-Kamchatkan family, consisting of Chukotian and Kamchadal (the latter represented by only one surviving language, Itelmen), but some writers (e.g. Worth 1962 and Volodin 1976) have not accepted this relationship. Beyond that there have been attempts to relate most of the binary combinations of the four US families: Uralic and Yukagir (Collinder 1940, Angere 1956, Tailleur 1959), Eskimo-Aleut and Chukchi-Kamchatkan (Swadesh 1962, Hamp 1976), Uralic and Eskimo-Aleut (Bergsland 1959), etc. These works are of varying quality, but most have been hampered by the paucity of good descriptive work on these languages (except for that on Uralic). The case for a Uralic-Yukagir connection has apparently been greatly strengthened recently by the appearance of Nikolaeva (1988), but I have been unable to obtain this work.

---

[6] This review by the late Peter A. Michalove (1951-2013) was apparently written several years ago, but never published, as far as we know. Thanks to Allan R. Bomhard for calling our attention to it. Ed.

Fortescue also mentions the possibility of relating all of these languages within a wider Nostratic or Eurasiatic context. He is sympathetic to the idea, but does not pursue it here because he feels that, even if there is a broader Nostratic configuration, the US languages form a more closely related subset. He feels that the Altaic languages, which have been those most frequently compared to Uralic, do not share most of the isoglosses that bind the US group together. However, I find Altaic correspondents to many of the morphological and lexical parallels he considers for the US languages. Further, he states that his view differs from most Nostratic work, "in not seeing any solid grounds for extending these 'stocks' further west into Europe or south into India" (p. 52).

Chapter three offers a typological overview of the languages of Siberia. While typological features are not generally considered reliable guides to possible genetic relationship in themselves, Fortescue makes ingenious use of typology in establishing historical connections among languages groups. Although some typological features, such as word-order and ergative or accusative alignment are extremely common, and others, such as postpositions and government relationships, are the result of implicational universals involving other features, Fortescue maintains that rarer features, when geographically clearly defined, can argue for language relationship. For example all of the US languages are characterized by a single (voiceless) set of plosives and a single (voiced) set of non-sibilant fricatives. This is an extremely rare combination among the languages of the world, and it is found nowhere in the languages surrounding US, including those with which they are known to have had extensive contact, such as Yeniseian, Altaic, Na-Dene, and Salishan.

Chapter Four summarizes the morphological evidence for a relationship between Chukchi-Kamchatkan and Eskimo-Aleut, while Chapter Five adds to the picture by bringing in Uralic and Yukagir. Significantly, Fortescue finds that the inclusion of the latter two only deepens the overall picture by confirming many of the commonalities discussed in Chapter Four. Building on the chapter on typology, many of these morphological commonalities are in striking contrast to the material and typological morphologies of surrounding languages.

Chapter Six deals with the lexical correspondences among these language families, and in so doing, sets out a reconstructed phonological system and proposed set of reflexes in the various languages. In this manner, Fortescue is able not only to present a large number of apparently cognate forms, often of quite different phonetic shape, but also to identify many phonetically common forms as loans because they do not conform to these phonological developments. One of the most striking aspects of his list of common US forms (defined as those represented in phonologically regular form in at least three of the four families) is the absence of word-initial */r-/, while initial */l-/ is well represented. This is a phonotactic characteristic of Tungusic as well, and most modem writers who accept the Altaic relationship reconstruct it for Altaic itself (e.g. Starostin 1991, Vovin 1994).

Chapter Seven discusses the archaeological evidence of settlement and migration in the far north, while the final chapter summarizes the cumulative historical picture suggested by the data. Fortescue sees an origin of the US languages in southern and central Siberia c. 8000 to 10,000 years BP.

The greatest weakness of this book is the very poor production quality. There is a huge number of typos, and errors in bibliographic references. The maps, especially those showing the distribution of isoglosses are poorly reproduced and difficult to read, and the endnotes after each chapter are less convenient then footnotes on the page would have been.

But this is a minor quibble. The possibility of a genetic relationship among these languages has long been discussed in mostly impressionistic terms. This book is one of the first efforts to lay out the evidence systematically, in a manner consistent with modem scholarly standards. As such, it is full of stimulating ideas, which are sure to be developed by further research. This book will be an important basis for that ongoing work.

References

Angere, Johannes. 1956. *Die uralo-jukagirische Frage: em Beitrag zum sprachlichen Urverwandtschaft.* Stockholm.

Bergsland, Knut. 1959. The Eskimo-Uralic hypothesis. *Journal de la Société finno-ougrienne* 61: 1-29.

Collinder, Björn. 1940. *Jukagirisch und Uralisch. Uppsala Universitets Årsskrift 1940*, 8.

Hamp, Eric. 1976. On Eskimo-Aleut and Louravetlan. *Papers on Eskimo and Aleut Linguistics*, ed. Eric Hamp, pp. 81-92. Chicago: Chicago Linguistic Society.

Nikolaeva, Irena A. 1988. *Problema uralo-jukagirskix genetičeskix svjazej.* Dissertacija kand. filol. nauk. Moscow: Akademija Nauk.

Starostin, S. A. 1991. *Altajskaja problema i proiskhoždenie japonskogo jazyka.* Moscow: Nauka.

Swadesh, Morris. 1962. Linguistic relations across Bering Strait. *American Anthropologist* 64: 1262-1291.

Tailleur, Olivier G. 1959. Plaidoyer pour le youkaghir, branche orientale de la famille ouralienne. *Lingua* 8: 403-423.

Volodin, Aleksandr P. 1976. *Itel'menskij jazyk.* Leningrad: Nauka.

Vovin, Alexander 1994. Long-distance relationships, reconstruction methodology, and the origins of Japanese. *Diachronica* XI.1: 95-114.

Worth, Dean. 1962. La place du kamtchedal parmi les langues soi-disant paléosiberiennes. *Orbis* 11:579:599.

News and Notices

Association for the Study of Language in Prehistory (ASLIP)

Mother Tongue XX (2015)

The next issue of *Mother Tongue* will be dedicated to the memory of its Founder, Harold C. ("Hal") Fleming. All ASLIP members and Mother Tongue readers are invited to contribute. Contributions may take the form of *personal tributes or memories*, or *scholarly articles*, or both.

Contributors may send submissions or inquiries to the Editor of *Mother Tongue*, J.D. Bengtson (see contact information on the inside front cover of this volume).

Annual Meeting

The annual meeting of ASLIP was held at 12:38 p.m., 9 November, 2014, at the Department of South Asian Studies, Harvard University, 1 Bow Street, Cambridge, Massachusetts, U.S.A. The meeting was attended by Michael Witzel (President), John D. Bengtson (Vice-President, *Mother Tongue* Editor), Michael T. Lewis (Secretary-Treasurer), Harold C. Fleming (ASLIP Founder, former President, and now on the Board of Directors), Nicholas Davidson (Board of Directors), B.K. Rana (Harvard University), and Caley Smith (Harvard University).

President Witzel reported that the tax-exempt status of ASLIP was now restored. It had lapsed, and its renewal was delayed by internal and external problems. ASLIP is now free to apply for foundation funding, and there was some discussion about conferences ASLIP could sponsor, for example:

- Conference discussing human **genetics** and/or **archeology** (e.g., Harvard's Reich & Meadow)
- Conference on the **Nostratic** hypothesis (Bomhard, "Moscow School," et al.)
- Conference on the **Dene-Caucasian** hypothesis (also "Dene-Yeniseian" & Swadesh's "Vasco-Dene")
- Conference on the **Khoisan** hypothesis: genetic family or areal convergence?

Funding is also needed for expanding the ASLIP homepage (http://aslip.org/) and for electronic publication of Mother Tongue.

ASLIP members are urged to contact the officers with ideas and suggestions about these and other possibilities (see contact information on the inside front cover of this volume).

The ASLIP Council of Fellows: "Council Fellow" is an honorary office, elected by ASLIP members as a whole. At the 2014 annual meeting the following were nominated for the Council:

- **Pierre J. Bancel:** Association d'études linguistiques et anthropologiques préhistoriques, Paris.
- **Günter Bräuer:** Universität Hamburg.
- **Larry Lepionka:** College of Charleston, South Carolina.
- **Shamil Nafiqoff:** Russian Academy of Sciences, Oufa, Bashkortostan, Russia.

ASLIP members are **urged to vote** for one or more of the candidates. Votes can be sent by snailmail or email to ASLIP Secretary-Treasurer, Michael T. Lewis (see contact information on the inside covers of this volume).

Mother Tongue Editor John D. Bengtson and Nicholas Davidson led a discussion about the possibility of publication being taken over by Gorgias Press, an academic publisher based in Piscataway, New Jersey. Such an arrangement could have benefits: it would relieve the Secretary-Treasurer of some cumbersome duties, and make *Mother Tongue* more available to the public, since issues would be available on online book outlets Amazon and Barnes & Noble. Mr. Davidson was appointed Administrative Editor of *Mother Tongue* and charged with handling negotiations with Gorgias Press.

Nicholas Davidson and Hal Fleming discussed *Anči*, the female figure that has graced every issue of *Mother Tongue* (Journal) and many issues of its newsletter predecessor. Hal explained that *Anči* represents the "eternal African woman," and by extension is the embodiment of "Mother Tongue," a.k.a. "Proto-Human" or "Proto-Sapiens," the putative language of "Mitochondrial Eve." (The word *anči* is literally the Amharic word for 'thou' [feminine]). It was mentioned that we ought to recover a sharper, more original form of the figure, since recent Mother Tongue issues have featured a copy of a copy (of a copy) of the original *Anči.*

* * *

Back Print Issues of MOTHER TONGUE

Back print issues of *Mother Tongue* are available for purchase. The following table summarizes some of the topics covered in issues I – XVII:

I (1995) **Inaugural Issue:** Canaanite & Bengali, Austric; Basque & Dene-Caucasian (R.L. Trask & 12 discussants); Proof in Genetic Linguistics (Greenberg)
II (1996): Kusunda, Ainu, Basque, Nihali (Mundlay & 8 discussants); Basque & Dene-Caucasian (S. Starostin, Trask, Ruhlen); Multilateral comparison (Greenberg)
III (1997): Kusunda, Nihali, Sumerian; "Hardware" / Origin of Language Symposium (Zegura, Lieberman, Donald, Fitch, Deacon); Recommendations for Long Rangers (Benedict); S.A. Starostin
IV (1998): Yeniseian; Ainu (Sidwell, Itabashi, Norquest, Bengtson); Deep classifications; Apophony (ablaut)
V (1999): Austric (Hayes, Blažek, Blust, van Driem, Fleming); Basque & Caucasian (Bengtson & 6 discussants); Sumerian (Srinivasan, Witzel, Diakonoff, Bengtson); Climatic influences on language; Bipeds, tools & speech; American prehistory
SPECIAL ISSUE (1999): **South Asian substrate languages** (Witzel, Whitehouse, van Driem, G.D.S. Anderson, Kuiper, Masica, Mundlay); Austronesian taxonomy
VI (2000/2001) **Festschrift for Roger W. Wescott**: Austric; Paleolinguistics: The State

of the Art and Science (10 discussants); Obituaries: Wescott, Gordon, Greenberg

VII (2002) **In Honor of Joseph H. Greenberg**: Elamite, Dravidian, Ongota, Shabo, Tasmanian, Andamanese, Eurasiatic; Greenberg's taxonomic proposals; Proto-Human or Proto-Sapiens

VIII (2003) Linguistic Databases & Taxonomy Workshop (SFI): Nostratic, Salishan & Caucasian, Basque, Khoisan, Negative Evidence (Whitehouse); EHL Project

IX (2004): Australian languages, Kadu, Ongota, Shabo; Australian languages (O'Grady & Whitehouse); Proto-Sapiens kinship words: (P)APA, (T)ATA; Mario Alinei

X (2005): Kusunda, Basque, Eurasiatic; Obituaries: Livingstone, S. Starostin, Greenberg; Flores "hobbits"; Great Archeological Debate; Pre-Clovis site; Chinese genome; Trombetti

XI (2006) **Asian Remnant Languages & Year of the Australoid** (Harvard / ASLIP Conference, 2006): Indo-Pacific, South Asian languages, Tibeto-Burman, Austroasiatic, Kusunda, Austric, Australian, Dravidian, Andamanese; Archeology of Southern Route (Harrod); Out of East Africa by 77K BP (Brooks); Population genetics

XII (2007) **In Honor of Harold C. Fleming's 80th Birthday**: Indo-European, Nostratic, Kartvelian, Bangi Me, Shompen, Dravidian; Nostratic Phonology (Bomhard, Sidwell, G. Starostin); Obituaries (Orel, Helimski, Bender); Glottochronology, Genetics

XIII (2008) **Commemoration of Ann Arbor Language & Prehistory Symposium (1988):** Milyan, Nostratic, Uralic, Chukcho-Kamchatkan, Shompen, Andamanese; Obituaries: Zvelebil, O'Grady; Bio-genetics; Fallacy of time limit; Myth of rapid linguistic change; Linguistic chronology

XIV (2009) **Commemoration of Daniel F. McCall**: Indo-European, Caucasian, Basque, El Molo, Mesmes, mystery languages of East Africa; Berber *H (Fournet, Blažek, Kossmann, Prasse); Paleoanthropology; Myth of rapid linguistic change II; Numerals (Hurrian, Nilotic); Profiles (Dolgopolsky, Mallory)

XV (2010) **Fifteenth Anniversary Issue 1995-2010:** Areal patterns of myth motifs (Berezkin); Holocene etymology of 'pitch'; Myth of rapid linguistic change III; Yeniseian numerals; Afrasian etymologies; Review of Campbell & Poser *Language Classification*

XVI (2011): Archeology & Genetics; Indo-European & Fenno-Ugric (Pedersen); Chinese giant Pangu; Minoan; Milyan; Surmic numerals; Dene-Caucasian; Myth of rapid linguistic change IV; Review of Jones & Milicic *Kinship, Language & Prehistory*

XVII (2012) **In Memory of Aharon Dolgopolsky:** Personal memories of Aharon Dolgopolsky; Archeology & Genetics; Trombetti's "Puluga" and discussion; Kamchukchean and Eskaleutian; Discussion: The Number 'One' (Levitt, Blažek, Bomhard, Bürgisser, Janhunen)

XVIII (2013) **Commemorating the 50th Anniversary of J.H. Greenberg's *The Languages of Africa*:** Obituaries: C.G. Turner, P.A. Michalove; Paris Linguistic Society Ban on discussion of language origins; The Niger-Congo hypothesis; Greenberg tribute and appraisal; Substrates and isolates of Africa; Proto-Nostratic morphology; The early dispersions of *Homo sapiens sapiens*; Origin of Milyan verbs; Global etymologies and Trombetti; Notes on Moscow conference (2000)

Back print issues can be had for $10 (domestic U.S.) or $20 (foreign). Please contact ASLIP Secretary-Treasurer **Michael T. Lewis**, 20 Duane Avenue, West Newton, MA 02465, U.S.A. | Tel. 617-964-0978 | lewismtc@rcn.com